

مقدمة

الأساليب الكمية في الجغرافيا

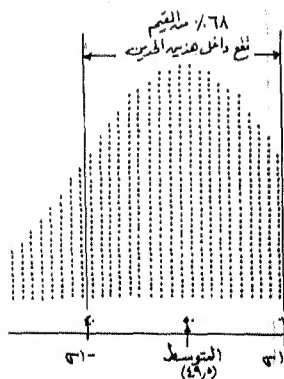
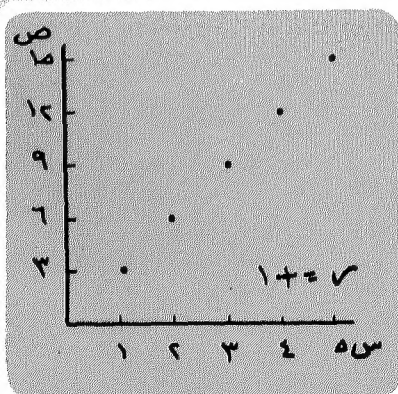
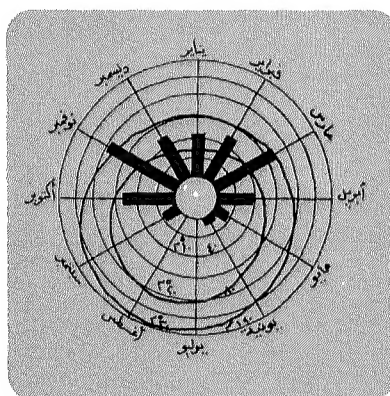
الأستاذ الدكتور

فتحي عبد العزيز أوراقي

أستاذ الجغرافيا الطبيعية

وعميد كلية الآداب، جامعة بيروت العربية

$$\frac{(1 - r^n) + 1 + (1 - r^n) + \dots + 1 + (1 - r^n)}{(1 - r^n) + \dots + (1 - r^n)}$$



Bibliotheca Alexandrina

0103221

مقدمة

الأساليب الكمية في الجغرافيا

دكتور

فتحى عبد العزيز أبوراضى

أستاذ الجغرافية الطبيعية

عميد كلية الآداب

جامعة بيروت العربية

٢٠٠٠

دار المعرفة الجامعية

٤٠ شارع سويتير - المنيا - جمهورية مصر العربية - ٤٨٣٠١٦٣
٣٨٧ شارع جمال الدين - القاهرة - ٥٩٧٣١٤٦

حقوق الطبع محفوظة

دار المعرفة الجامعية

للطباعة والنشر والتوزيع

الإدارة : ٤٠ شارع سوتير



الأزاريطة - الاسكندرية

ت : ٤٨٣٠١٦٣

الفرع : ٣٨٧ شارع قنال السويس



الشاطبي - الاسكندرية

ت : ٥٩٧٣١٤٦

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

إهداء

إلى زوجتي:

شريكة الحياة

وصديقة العمر

التصدير

شهدت سنوات العقود الثلاثة الماضية تغييراً كبيراً وتطوراً ملحوظاً في علم الجغرافيا، ليس في منهجه ومحتواه فحسب، وإنما أيضاً في الأساليب التي يعتمد عليها في تحقيق أغراضه وأهدافه. وليس من شك في أن أهم التطورات التي شهدتها هذا العلم هو التعامل مع الأرقام أو ما عرف بالانجاء الكمي المتمثل في تطبيق الأساليب الاحصائية الكمية في تحليل العلاقات المختلفة بين مكونات البيئة ونشاط الإنسان، وفي دراسة المشكلات والظواهر الجغرافية المتنوعة سواء كانت طبيعية أو بشرية بغية الوصول إلى نتائج رقمية محدودة تختصر كثيراً من التحليلات الوصفية الكيفية (النوعية) في تشخيص وتفسير الظواهر الجغرافية.

ورغم أن لغة الأرقام ليست بغريبة على بعض فروع علم الجغرافيا، إذ تعد البيانات الرقمية مصدراً رئيسياً تعتمد عليه مجالات التحليل وأساساً للبحث والدراسة منذ البداية في فروع مثل الجغرافية الاقتصادية وجغرافية السكان والجغرافية المناخية، فإن ما شهدته السنوات الأخيرة يعد محاولة للتعمق في التعامل مع الأرقام وتحولاً نحو استخدام الأساليب الكمية في التحليل الرقمي في الدراسات والبحوث الجغرافية بعامه. وكان لهذا التطور أو التحول نتائج هامة أسفرت عن دفع عجلة هذا العلم والارتقاء به إلى مصاف العلوم الأرضية والاجتماعية التي اتخذت لنفسها

مساراً ديناميكياً كمياً. ولذا فقد أطلق البعض على هذا التطور أو التحول في استخدام أسلوب التحليل الرقمي في الجغرافيا إسم الثورة الكمية (Burton, 1963) والثورة المفاهيمية (Davis, 1972). وقد لقي ذلك ترحيباً من بعض الجغرافيين، وتحمساً، وأحياناً تعصباً، من البعض الآخر، ولكن بعد أن هدأت «الثورة» ظهر إسم «الجغرافية الكمية Quantitative Geography كمصطلح شامل لكل ما حدث من تطور وتحول في الجغرافيا، ولذا فلا غرو أن يكون هو المصطلح السائد بين الجغرافيين المعاصرين. والجغرافية الكمية إذن ليست فرعاً من الفروع المعروفة للجغرافيا، كما أنها ليست فرعاً «جديداً» New Geography، وإنما هي تشير إلى واقع جديد أخذ يفرض نفسه على الجغرافيا، أو بعبارة أخرى هي اتجاه جديد في معالجة المشكلات الجغرافية بعامة. ويتمثل هذا الواقع أو الاتجاه في مجموعة الأساليب الإحصائية الكمية التي كان لكثرة تطبيقها واتساع مجال استخدامها وتعدد الدراسات فيها أن أدخلت قواعد وقوانين جديدة على الفكر الجغرافي، كما مهدت اكتشاف آفاق جديدة لميدان الدراسات الجغرافية. وكان من نتيجة ذلك أن تزايد الاهتمام الكمي وتطور في مختلف الفروع الجغرافية، وليس أدل على ذلك من أنه لانتحوا الآن المناهج الدراسية للجغرافيا في أقسام الجغرافيا بجامعة العالم المتقدم من منهج في «الجغرافية الكمية».

ويهدف هذا الكتاب إلى محاولة تطبيق بعض الأساليب الكمية في الجغرافيا واستخدامها كأدوات لحل المشكلات واستخلاص النتائج لاتخاذ القرارات التي هي أساس وهدف البحث العلمي النهائي. ومعنى ذلك أن هذا الكتاب لن يتناول دراسة وتطبيق كل الأساليب الممكنة في هذا المجال بل سيقترن على «عينه» من «مجتمع» الأساليب الكمية الذي مازالت الدراسة فيه ميداناً متسعاً تستقي منه كثير من الموضوعات في مجال الدراسات الكمية المتقدمة.

ولقد أعدت موضوعات هذا الكتاب، التي هي حصيلة جهد وخلاصة تجربة تدريس لهذا اللون من الدراسة على امتداد نحو ست عشرة سنة لطلاب الجغرافيا في جامعات شفيلد ونوتنجهام - إنجلترا - والاسكندرية. لكي تكون بمثابة مقدمة (مدخل) لدراسة الأساليب الكمية في الجغرافيا، تستطيع أن تعطى الباحث المبتدئ

معرفة منظّمة عن الأساليب والأدوات المستخدمة في جمع وتحليل بيانات الظاهرة الجغرافية، كما يمكن أن تكون أساساً يعتمد عليه الباحث المتخصص في تحليله للحقائق الرقبة الخاصة بالمشاكل التي تتعرض لها دراسته كبنية الوصول إلى أقصى ما يمكن من نتائج عليه مفيدة. ولقد حرصنا في ضوء هذا الهدف أن يكون عرض هذه الأساليب مبسطاً ولكنه وافياً وشاملاً للعديد من المبادئ التي تعد ضرورية وأساسية لمعالجة المشكلات الجغرافية. كما حاولنا بقدر المستطاع أن نتجنب الإثباتات والبراهين الرياضية حتى يتمكن الباحث الذي لا يتمتع بأية خلفية رياضية من فهم واستيعاب هذه الأساليب ومتابعة تطبيقها دون عناء. وأياً كان الأمر فإن اختيار وعرض موضوعات مثل هذا الكتاب ليس بالأمر اليسير، إذ لا توجد «طريقة» معروفة أو متفق عليها لتقسيم وتنظيم موضوعات هذا اللون من الدراسة، ولذا فإن التسلسل في موضوعات (فصول) هذا الكتاب يتخذ من المشكلات المختلفة التي تواجه الباحث الجغرافي أساساً له مثل تلخيص البيانات والمعطيات الخاصة بموضع (حيز) واحد في مكان (مجال أو فراغ) معين، أو توضيح المقارنات والعلاقات بين متغيرين أو أكثر. أو بمعنى آخر موضوع الكتاب بعامة يتجه نحو تحليل ومعالجة المشكلات الجغرافية أكثر من اتجاهه نحو دراسة الأساليب المستخدمة لذاتها وذلك على الرغم مما يوحى به أو يشير إليه، عنوان الكتاب «الأساليب الكمية». وحتى لا يحدث بعض من اللبس فإننا قمنا بتخليص الخصائص الأساسية والأهداف الرئيسية المنشودة من الأسلوب المستخدم في التحليل في بعض المواضع، إذا لزم الأمر، بالإضافة إلى شرح أمثلة تطبيقية لمشكلات نوعية توافق to fit خصائص وأهداف الأسلوب المستخدم في مواضع أخرى.

ويتكون إطار هذا الكتاب من مقدمة وأربعة أبواب تضم اثنتا عشر فصلاً وتتناول المقدمة شرح المفاهيم الاحصائية التي ستردد ذكرها في كثير من المواضع، وبيان الخصائص الرئيسية التي ينبغي توافرها في البيانات الاحصائية التي تمثل المادة الخام لأسلوب التحليل الكمي. وعرضنا في نهاية هذه المقدمة لموضوع الدقة والأخطاء في البيانات وأسبابها. ويعرض الباب الأول والذي اشتمل على الفصول الثلاثة الأولى لطرق جمع البيانات وإعدادها للتحليل الكمي. وقد اشتمل هذا الباب على مناقشة تحليلية مختصرة للمبادئ الأساسية لأساليب تحديد حجم العينات المتعارف

عليها، مع ربط هذه المبادئ بالواقع الجغرافى. وقد قمنا فى هذا الباب أيضاً بدراسة طرق العرض البيانى والجدولى كمنطلق أساسى لتحليل الاحصائى الكمى.

ويختص الباب الثانى يتكون من ثلاثة فصول، من الفصل الرابع حتى الفصل السادس، بمناقشة موضوع الوصف الكمى الذى يعتبر عملاً أساسياً فى كل العلوم - بما فيها الجغرافيا - عن طريق استخدام مقاييس محددة ومعايير ثابتة متفق عليها، وقد عرضنا فى هذا الباب مقاييس النزعة المركزية (المتوسطات) وبيننا كيفية تطبيقها والفروق فى النتائج المستخلصة من كل منها. وبعد ذلك انتقلنا إلى مناقشة مقاييس تجانس واختلاف (تشتت) البيانات وانحرافها (تباعدها) عن قيمة أحد مقاييس المتوسطات وبيننا كيفية استخدامهما وخصائصهما ومجال تطبيقاتهما المتعددة إلى جانب عرض مفصل لمزايا استخدام المقاييس ومشاكل تطبيقاتها، وقد تناولنا فى هذا الباب أيضاً موضوع الاتجاه وشكل تركيز بيانات المتغيرات فى أحد نواحي توزيعها، أو ما يعرف بمؤشرات التركيز، إذ لا تكفى مقاييس المتوسطات والاختلاف (التشتت) فى وصف وتشخيص التوزيعات ومقارنتها بعضها البعض لتحديد خصائصها وملامحها. ويشتمل الباب الثالث على أربعة فصول، من السابع حتى العاشر، تختص باستعراض التقدير الاحصائى وأساليب المقارنة بين مجموعات البيانات. وقد ناقشنا وبيننا طرق استنتاج وتقدير خصائص المجتمع من البيانات التى جمعت بأسلوب المعاينة بسبب عدم إمكانية دراسة وفحص جميع مفردات المجتمع التى جمعت بأسلوب المعاينة بسبب عدم إمكانية دراسة وفحص جميع مفردات المجتمع التى قد تكون كثيرة جداً. وانتقلنا بالمناقشة فى هذا الباب إلى قواعد اختيار الفروض الاحصائية التى تعتمد عليها أساليب المقارنة بين البيانات. وبيننا أساليب معالجة البيانات ومقارنتها فى صورة عملية تطبيقية على البيانات الجغرافية للوقوف على أهمية الأساليب العلمية (البارامترية) والأساليب غير العلمية (غير البارامترية) التى تستخدم هذا الشأن.

ويخص الباب الأخير من الكتاب وهو الباب الرابع باستعراض وتحليل العلاقات بين المتغيرات وتحديد اتجاهها العام. ويشتمل هذا الباب على ثلاثة فصول من الحادى عشر حتى الثالث عشر. وقد خصص الفصل الأول منها لتحليل الارتباط

لمعرفة درجة واتجاه العلاقة بين المتغيرات. والفصل الثانى من هذا الباب يعالج تحليل الانحدار لتمثيل العلاقة بين بيانات متغيرين بطريقة رياضية يمكن التنبؤ بها والحصول منها على بيانات متغير معين كلما تغير الآخر. وبهذا يمكن تقدير سلوك أحد المتغيرات فى ضوء تأثيره بمتغير آخر أو بعدة متغيرات أخرى.

والفصل الثالث من هذا الباب يتناول بالشرح والتفصيل تحديد الاتجاه العام الذى يعكس تأثير العوامل المختلفة التى تؤدى إلى التغيرات أو التطورات فى الظواهر كمياً عبر الزمن. وقد بينا أن استمرار الاتجاه العام فى المستقبل هو الأساس المنطقي للتنبؤ الاحصائي الذى يعتمد عليه عند اتخاذ القرارات فيما يتعلق بتخطيط المستقبل، والهدف الرئيسى لتحليل «السلال الزمنية» عن طريق دراسة مركباتها الأساسية وأهميتها فى استنتاج التذبذبات وتبعتها فى منحنى الظاهرة.

وحتى تكتمل الفائدة العلمية من موضوعات الكتاب، فقد حرصنا على تزويده بأمثلة عديدة وتطبيقات كثيرة لبيانات جغرافية لتوضيح استخدام وتطبيق أدوات وأساليب التحليل الاحصائي الكمي وتفسير نتائجها، بالإضافة إلى الكثير من الأشكال والرسوم البيانية حتى تكتمل متابعة الحقائق الواردة فى متن الموضوعات، كما لم يفتنا إلحاق مجموعة من الجداول الاحصائية التى يستعان بها فى عملية المعايرة الاحصائية.

ولاندعى أن الكتاب يخلو من نقائص فليس فى وسع أى باحث مهما كانت مقدراته العلمية أن يصل بدراسة إلى درجة الكمال - فهو لله وحده - ولكنها محاولة أرجو من خلالها أن أكون قد حققت ولو إضافة بسيطة إلى المكتبة العربية التى تعاني النقص الشديد من مؤلفات فى الأساليب الكمية فى مجال البحوث والدراسات الجغرافية بخاصة. كما أن كل أملى أن تحقق هذه المحاولة الهدف المنشود منها، وتسد هى وأمثالها الفجوة العلمية التى تفصل بين البحث الجغرافى فى الجامعات العربية وبقية جامعات العالم المتقدم، وأن يكون «تباين» المعروض من الموضوعات واتساع وتعمق الدراسة فيه خير معين للباحثين والمهتمين بالدراسات الجغرافية المتطورة، ومشجع لهم على تفهم طبيعة وخصائص الأساليب الكمية والتعامل معها فى بحوثهم ودراساتهم ففى ذلك مواصلة السير فى نهج المعرفة المتطورة ومواكبة التقدم العلمى الخلاق.

وأود بهذه المناسبة أن أتقدم بالشكر الجزيل لأستاذي الجليل الأستاذ الدكتور
جودة حسنين جودة الذي كان لتوجيهاته السديدة القيمة ولتشجيعه المتواصل لى
أثر كبير دفعنى نحو إصدار هذا الكتاب. كما أتوجه بجزيل الشكر وعظيم الامتنان
إلى أساتذتى الأفاضل وزملائى فى أقسام الجغرافيا بجامعة الإسكندرية وجامعتى
شفيلد ونوتنجهام بانجلترا على ما قدموه لى من عن صادق وتشجيع دائم خلال
مرحلة دراستى وتدريسى لهذا الموضوع، والذين أفدت كثيراً من توجيهاتهم
وملاحظاتهم أثناء مرحلة إعداد هذا الكتاب.

ويبقى أن أرجو بعد العناء أن أكون قد وفقت إلى أن أوفى فيما أقصد إليه على
غاية، وأن يحقق هذا الكتاب الغرض من إصداره.

والله الموفق والمستعان ،،

دكتور فتحى عبد العزيز أبراضى

الإسكندرية فى ١٠ / ٨ / ١٩٩٩

مقدمة
في
المفاهيم الإحصائية

مقدمة فى المفاهيم الإحصائية

تحتل - الآن - أساليب التحليل الإحصائي الكمي أو «الطرق الإحصائية» كأساليب علمية وأدوات بحث، أهمية خاصة فى الأبحاث العلمية الحديثة؛ إذ لا يكاد يخلو أى بحث من دراسة تحليلية إحصائية تتعرض لأصل الظاهرة قيد البحث فتصور واقعها فى قالب قياسى رقمى وتنتهى إلى إبراز اتجاهاتها وعلاقتها بالظواهر الأخرى، كما لا توجد أى دراسة دون أن تكون قد اتخذت لنفسها - أو لأحد جوانبها على الأقل - مناراً كمياً قائماً على الأساليب الإحصائية والرياضية الحديثة ونظرياتها المتجددة ونماذجها المتطورة، وعلى أحدث الآلات الحاسبة وأسرعها. ولا يقتصر هذا الاتجاه، الذى يشير إلى واقع جديد وتطور حديث، فى طريقة دراسة المشكلات ومعالجة البيانات على فرع من فروع العلم بل تتبناه جميع الفروع تقريباً للاستفادة منه، حتى أصبح استخدام الإحصاء أو الطرق الإحصائية سمة رئيسية وضرورة ملحة للبحث العلمى.

ويشير مصطلح «إحصاء» إلى معنيين رئيسيين (Hays, 1970): فهو فى معناه الضيق يستخدم للتعبير عن الكميات الهائلة من البيانات Data التى تجمع عن

طريق الإستفتاء أو التجارب أو الحصر مثل الإحصاء السكاني، كميات الإنتاج الزراعى والصناعى، وحجم التبادل التجارى بين الدول. أى أنه بهذا الإستخدام يختص بالحقائق والأرقام Facts and Figures. أما المعنى الثانى فيختص بالطرق العملية والأساليب العلمية (البنية والمعقدة) فى معالجة وتحليل وتفسير البيانات بغرض الوصول إلى نتائج مقبولة وقرارات سليمة. والإحصاء بمعناه الأخير يمكن اعتباره أحد فروع الرياضة التطبيقية بما يحمل بين طياته من نظريات وقوانين تساهم بدرجة كبيرة فى اتخاذ القرارات التى أصبحت الأساس والهدف النهائى للبحث العلمى الخلاق.

والإحصاء كعلم يعتبر من العلوم الحديثة نسبياً. إذ بدأ يطل على العلوم الأخرى ويتصل بها فى أواخر القرن التاسع عشر. ولو أن لهذا الاتصال جذور أو ربما بدور وضعت خلال القرنين السابع عشر والثامن عشر حين اهتم علماء الرياضيات بوضع نظرية الاحتمالات (Hammond and McCullagh, 1974). ففى سنة ١٧٧٠ استخدم مصطلح «الإحصاء Statistics» لأول مرة، ولكن بمعنى يختلف عن معناه فى الوقت الحاضر. فقد احتوى كتاب هوپر Hooper الذى نشر فى عام ١٧٧٠ بعنوان: "The Elements of Universal Erudition" فصلاً بعنوان «الإحصاء، Statistics» يتضمن تعريفاً للإحصاء على أنه أحد فروع علم السياسة الذى يهتم بجمع وتصنيف وتحليل الحقائق التى لها تأثير على حالة الدول المعروفة فى ذلك الوقت (Yale and Kendall, 1953). وفى أوائل القرن التاسع عشر حدث تغير (تحول) فى مفهوم علم الإحصاء، فأصبح يعرف على أنه علم «الحساب السياسى» الذى يستخدم الطرق العددية Numerical Methods فى إبراز الشخصية الاقتصادية للدولة. ونهاية القرن الماضى استخدمت الطرق الإحصائية كأداة لتلخيص البيانات ووصفها فقط. ثم زاد الاهتمام بعلم الإحصاء مع بداية القرن الحالى حتى أصبح الآن علماً له قواعده ونظرياته، وأصبحت الأساليب الإحصائية تطبق كوسيلة لاستنتاج الحقائق فى شتى فروع العلوم الأرضية Earth Sciences

سواء كان فى العلوم الطبيعية كالكيمياء والطبيعة أو فى العلوم الاجتماعية كالجغرافيا والاجتماع وعلم النفس وغيرها.

ونظراً لأن مجال الدراسة التى بين أيدينا لاتمكننا من التعرف - بالتفصيل والتدقيق - على التطور التاريخى لعلم الإحصاء، تبعاً لتعدد طرق ومداخل دراسة هذا التطور، إلا أنه قد يكون من المفيد والمناسب هنا الإلمام بخواص هذا العلم عن طريق عرض مختصر لأهم وظائفه واستخداماته والمفاهيم الخاصة به.

الوصف الاحصائى Statistical Description

تعد وظيفة الوصف من الوظائف الأولية لعلم الإحصاء التى تستخدم فى تلمس حقائق الظواهر المختلفة (جغرافية، اجتماعية، اقتصادية ... إلخ). ويعتمد أسلوب الوصف الإحصائى على المقاييس والمؤشرات الإحصائية، بعضها خاص بقياس القيم المركزية للبيانات. والبعض الآخر خاص بقياس مدى دقة هذه المقاييس. بغرض تقصى الحقائق وتحديد الخصائص العامة دون الوصول إلى نتائج أو استدلالات. وباستخدام هذا الأسلوب فى وصف الظواهر بعد تلخيصها وعرضها على هيئة جداول أو أشكال بيانية يصبح من السهل تحديد خصائصها واتجاهاتها العامة بطريقة علمية منظمة.

الاستدلال (الاستنتاج) الاحصائى Statistical Inference

يعرف الاستدلال الإحصائى بأنه عبارة عن التوصيات أو الاستنتاجات المبنية على طرق إحصائية تتناسب مع ظروف الظاهرة موضع الدراسة. وتتناول تحليل بياناتها. ويمكن تقسيم الاستدلال الإحصائى إلى نوعين رئيسيين (Spiegel, 1972) النوع الأول يطلق عليه الاستدلال الإستنتاجى Deductive Inerence، وهو عبارة عن تطبيق تفسير ظاهرة خاصة من نتائج دراسة ظاهرة عامة. فعلى سبيل المثال إذا ذكرنا أن انخفاض نسبة وفيات الأطفال عموماً يرجع إلى ارتفاع المستوى الصحى

(ظاهرة عامة)، فإنه يمكن القول أن ارتفاع المستوى الصحي لسكان منطقة ما سيؤدى، بالضرورة، إلى انخفاض نسبة وفيات الأطفال بها. أما النوع الثانى من الاستدلال فهو الاستدلال الإستقرائى Inductive Inference، وهو تطبيق تفسير ظاهرة خاصة على ظاهرة عامة. وهذا النوع من الاستدلال يستخدم فى إطار التجارب المعملية، حيث أنه يمكن التحكم فى جميع الظروف المحيطة والمؤثرة فى الظاهرة موضع الدراسة. ولكن تحت الظروف الطبيعية فإن أى ظاهرة عرضة للاختلاف والتباين تبعاً لتأثير العوامل البيئية المتعددة، وعليه فلا يكون الاستدلال مؤكداً، وحيث ما هو صحيح بالنسبة للظاهرة الخاصة قد لا يكون أن صحيحاً بالنسبة للظاهرة العامة، فإن لغة الاحتمال تستخدم عند عرض النتائج والتوصيات. فمثلاً تفسير ظاهرة التعرية الساحلية فى منطقة ما لا ينطبق بالضرورة على نفس الظاهرة فى كل جهات العالم الساحلية. إذ أن العوامل الطبيعية التى تسبب هذه الظاهرة تختلف من منطقة إلى أخرى حسب الظروف البيئية السائدة فى كل منطقة. ولكن يمكن القول بأنه من المحتمل أن تشترك جهات العالم الساحلية التى تعاني من ظاهرة التعرية فى بعض العوامل المسببة. أى أن الاستقراء يعبر عنه فى هذه الحالة فى صورة احتمالية.

اختبار الفروض الاحصائية Testing of Hypotheses

عند تحليل البيانات الاحصائية يقوم الباحث بتلخيص الغرض من الدراسة وذلك على هيئة فرض مقترح أو نظرية، تتناسب مع خواص متغيرات ظاهرة، يمكن اختبارها والحكم على صلاحيتها لتفسير مدى إنطباق الفرض الموضوع على النتائج المتحصل عليها. فإذا كانت النتائج تتفق مع الفرض المقترح فإننا نقبله، وبالتالي يمكن تعميمه. أما إذا ظهر تعارض بين النتائج المستخلصة والفرض المقترح، فإن هذا الفرض سوف يرفض. ويتم قبول أو رفض الفرض المقترح

باستخدام الأساليب الإحصائية التي يتم بواسطتها تحليل البيانات الإحصائية للظاهرة موضع الدراسة، والتي عن طريقها يمكن الوصول إلى النتيجة العملية والقرار المناسب في هذا الشأن.

ويتم الأسلوب الإحصائي لاختبار الفروض Testing of Hypotheses من خلال المشاهدة المتكررة للتغير في الظاهرة، وعلاقة هذا التغير بالفرض المقترح، أو ما يسمى إحصائياً بفرض العدم Null Hypothesis. فإذا ما توصلنا إلى عدم وجود فرق جوهري أو حقيقي بين المشاهدات وما تم افتراضه، فإن الفرض المقترح يكون صحيحاً إحصائياً في حدود خطأ مسموح به عند مستوى دلالة معين. وفي حالة توصلنا إلى وجود فرق حقيقي (معنوي) بين ما تم قياسه من واقع المشاهدات وما تم افتراضه فإن الفرق يكون غير صحيح، لأن المشاهدات الواقعية لا تؤيد ما كان يتوقع في تغير بيانات الظاهرة موضع التحليل.

وتعتبر الاختبارات الإحصائية للفروض بمثابة الأسلوب العلمي في استخلاص النتائج بطريقة موضوعية دقيقة. كما أنها مفيدة جداً في شرح وتوضيح بعض الحقائق عن طبيعة البيانات الإحصائية.

التنبؤ (التوقع) الإحصائي: Statistical Prediction

يقصد بالتنبؤ، كمفهوم إحصائي، هو تلك التغيرات التي حدثت لظاهرة ما في الماضي، وليس في المستقبل. وذلك لتأكيد وجود الظاهرة من خلال المشاهدة والقياس، واختبار الفروض وتفسير التغيرات واستخلاص النتائج. وتعتمد دقة التنبؤ اعتماداً يكاد كلياً على مبدأ «الاحتمالية Determinism» في الظاهرة موضع التنبؤ. والذي يؤدي إلى استخلاص نتائج متشابهة تحت ظروف متشابهة. ولنضرب مثلاً على ذلك بالجاذبية الأرضية. فمن المعروف أن سرعة أى جسم في الفراغ ترجع إلى الجاذبية نحو الأرض التي قدرت بنحو ٩,٧٥ متر/ ثانية/ ثانية. وبواسطة هذا القانون يمكن لنا أن نحسب - أو نتنبأ بتأكيد تام - طول المسافة التي سيقطعها جسم ساقط في وقت معلوم، أو بعبارة أخرى يمكن التنبؤ بسرعة هذا الجسم في

لحظة معلومة خلال فترة سقوطه. وفي الجغرافيا لا نجد سوى عدداً قليلاً من الظواهر الجغرافية التي تتصف بالطبيعة الحتمية، أما الغالبية العظمى منها فتتصف بتأثر بعضها البعض بطرق متباينة، وفي أوقات مختلفة. فنادر ما يمكن الحصول على نتائج نهائية في دراسة أى منها حتى ولو كان ذلك تحت ظروف معينة، أو بوضع شروط أو فروض محددة.

والتنبؤ Prediction بمفهومه الاستدلالي السابق هو تنبؤ يخص الماضي Post-diction، وهو يختلف عن تنبؤ المستقبل Forecasting الذى يستخدم فيه التحليل الإحصائي للتوصل إلى توضيح الاتجاه العام لما سيحدث فى المستقبل للمتغيرات التى تتحكم فى تطور ظاهرة ما. وكذلك يبان العلاقات بين متغيرات الظاهرة لفترة مستقبلية.

المجتمع والعينة Population and Sample

يعرف المجتمع الإحصائي Population أحياناً باسم «العشيرة» وهو يتكون من جميع الأفراد Individuals (القيم Values) موضع الاستقصاء، والمطلوب معرفة خصائصها وتحديد الحقائق عنها. مثال ذلك المجتمع السكاني فى دولة ما، والمجتمع الحيوانى والنباتى، وكذلك مجتمع المصانع أو مجتمع الوحدات المنتجة من مصانع سلعة معينة.

والمجتمع الإحصائي إما أن يكون محدوداً Finite، أى أنه يمكن حصر جميع أفرادها مثال ذلك مجتمع السكان فى مدينة ما فى سن معين (١ - ٥ سنوات). أو إنتاج مصنع للسجاد فى يوم معين. وإما أن يكون غير محدود Infinite، ومن أمثلته مجتمع النقاط التى يتكون منها خط، أو عدد القياسات لمساحة ما، أو المجتمع المكون من جميع النتائج الممكنة (صورة، كتابة) فى قذفات متتالية لعملة معدنية. وعلى العموم فإن المجتمع الإحصائي يعد غير محدود فى حالة صعوبة حصر أو قياس الأفراد التى يتكون منها. وعند دراسة المجتمعات الإحصائية لا يجب تطبيق القوانين الخاصة بالمجتمعات غير المحدودة على المجتمعات المحدودة. إلا إذا كانت الأخيرة مكونة من أفراد يزيد عددها عن عدة مئات أو آلاف.

وتعرف المقاييس الإحصائية الخاصة بالمجتمع والمميز له باسم معالم (ثوابت)

المجتمع Parameters مثل المتوسطات والانحراف المعياري للمجتمع. وكما ذكرنا، فإنه قد يكون من المستحيل حساب هذه المعالم من المجتمع لصعوبة حصر جميع مفرداته، إلا أنه يمكن أن نستخلصها عادة من عينة مأخوذة من نفس المجتمع.

والعينة Sample هي جزء صغير من مفردات المجتمع الإحصائي تؤخذ لتمثل المجتمع، وتدرس بهدف الحصول على نتائج مهمة عن المجتمع عن طريق تحليل بياناتها. لذلك يجب أن تكون العينة ممثلة تمثيلاً صحيحاً للمجتمع الذي أخذت منه.

وتعرف المقاييس المحسوبة من بيانات العينة باسم «إحصائيات» أو مقاييس العينة Statistics وهي مقاييس تقديرية Estimate لما يقابلها من «معالم» المجتمع التي تمثلها.

البيانات (المعطيات) الإحصائية Data

يقصد بتعبير البيانات أى «كمية من المعلومات فى صورة رقمية». والصورة الرقمية للبيانات تبدو إما على شكل أرقام صحيحة Integers مثل ١٠، ١١٢، ٤٦٤... إلخ، أو على شكل أرقام حقيقية Real Numters مثل ٤، ٢٠، ٨، ٦١، ١٨٢، ٠، ٣١٧ أى أنها الأرقام التى تحتوى على علامة عشرية.

وتعد المعلومات الرقمية (البيانات الكمية) المادة الخام لأسلوب العمل الإحصائي، كما أنها تلعب دوراً كبيراً فى تطبيق أساليب التحليل الإحصائي الكمي حيث أن هذا النوع من البيانات يمكن قياسه كمياً مما يسهل استخدام هذه الأساليب لاستخلاص النتائج واتخاذ القرارات. وهناك بعض المصطلحات الخاصة بالبيانات الإحصائية التى سوف ترد كثيراً فى متن الفصول القادمة، نعرضها بصورة مختصرة قبل الخوض فى دراسة أساليب التحليل الإحصائي.

المفردات والمتغيرات Individuals and Variables

المفردة فى الإحصاء عبارة عن وحدة قياس المجتمع الإحصائي، ويعبر عن المفردات فى البيانات الإحصائية بالتمييز العددي للأشخاص كأعداد الطلبة والأسر والعمال، أو للحيوان مثل عدد الأبقار وعدد الأغنام، أو للجماهير مثل عدد المصانع وعدد المدارس وعدد المستشفيات.

والمغيرات Variables عبارة عن ظاهرات أو صفات تختلف قيمهما باختلاف الحالات Cases. والقيمة التي تعطى لكل مفردة من صفة معينة تعرف باسم Vari-ate. ويطلق على الصفات التي تتغير عشوائياً اسم المتغيرات العشوائية Random Variables وهي تلك التي تحدث بالصدفة مثل درجة الحرارة وشدة سقوط الأمطار. وتنقسم المتغيرات في قيمها العددية إلى قسمين هما: المتغيرات المتصلة (المستمرة) Continuous Variables وهي المتغيرات التي يمكن أن تأخذ أى قيمة (أعداد صحيحة وكسور) على المقياس المستخدم. فمثلاً إذا إرتفعت درجة الحرارة يوم من ٢٠ درجة مئوية إلى ٣٠ درجة مئوية خلال الترمومتر الزئبقى، فمعنى ذلك أن الزئبق يكون قد مر بكل القيم الواقعة بين هاتين الدرجتين. كذلك الحال فى مقياس سرعة السيارة، فإذا زادت السرعة من ٣٠ كيلو متر/ ساعة إلى ٦٠ كيلو متر/ ساعة فإن المؤشر فى المقياس يكون قد مر على كل القيم المحصورة بين هذين الرقمين. وبالمثل أيضاً الأطوال، وذلك لأن طول الشخص قد يكون ١٦٨ سنتيمتر أو ١٦٨,١ أو أى قيمة مهما كانت كسرية، وأصغر من المليمتر إذا كان المقياس يسمح بذلك. والنوع الآخر من المتغيرات يطلق عليه المتغيرات غير المتصلة أو الوثابة Discrete Variables وهي التي تختلف قيمها من مرحلة إلى أخرى بدون أن تكون منتظمة، كما أن قيمها لا تأخذ إلا أعداداً صحيحة Integrals. فعدد الرحلات التي يقوم بها الأشخاص، وكمية الفيضان فى الأودية الصحراوية، وعدد الأنهار والبحار والقارات، وعدد السيارات المارة فى أحد الشوارع، وعدد الفصول فى المدارس وعدد الحجرات بالمنازل، وحجم الأسرات ... إلخ، كلها متغيرات وثابة (غير متصلة) نحصل عليها فى الغالب بالعدد، ولذا تعرف بيانات هذه المتغيرات أحياناً بالبيانات العددية Enumeration Data.

أنواع البيانات:

تمثل الطرق المختلفة التي تقاس بواسطتها البيانات أهمية خاصة لأساليب التحليل الإحصائي الكمي، إذ أن لكل أسلوب منها طريقة خاصة تقاس أو تجمع على أساسها البيانات. وبشكل عام هناك أربع طرق مختلفة تقاس بواسطتها أربع مجموعات رئيسية من البيانات هي: البيانات الإسمية، البيانات الترتيبية، بيانات الفترة، وبيانات النسبة.

(١) البيانات الإسمية (النوعية) : Nominal Data

تشتمل قياسات خصائص الظاهرة موضع الدراسة فى هذا النوع من البيانات على قياسات ثنائية أو ثلاثية. ولنضرب مثلاً على ذلك، فعند تسجيل الحالة التعليمية لمجموعة من الأشخاص، فإننا فى هذه الحالة نعطى للشخص الأسمى الرقم (١) والشخص المتعلم الرقم (٢)، وإذا كانت الدراسة تتعلق بتسجيل نوع التربة: طينية أم رملية فى منطقة ما، نعطى التربة الطينية الرقم (١) والتربة الرملية الرقم (٢) وإذا لم تكن التربة طينية أو رملية تعطى الرقم (صفر).

ويطلق على البيانات الإسمية أحياناً اسم البيانات التصنيفية: لأنها تصنف المتغيرات على أساس خصائصها. ومن أمثلتها تصنيف التربة إلى التربة البنية، تربة البودزول، تربة اللاتريت ... إلخ، وتصنيف النباتات الطبيعية إلى غابات، حشائش وصحارى. وكلها أمثلة لا تحمل الصورة الكمية. وتعد طريقة قياس البيانات الإسمية من أضعف مستويات القياس لإستخدامها الأسماء أو الترتيم أو الرموز. وإستخدام الرموز فى مثل هذا النوع من البيانات ليس له أى علاقة بنوع الظاهرة أو درجة تركزها. فمثلاً فى وصف لون الحجر الرملى يمكن أن نوضع بيانات الوصف كالآتى:

صفه اللون :	أحمر	رمادى	أصفر	أبيض
أو (بالرقم) :	١	٢	٣	٤
أو (بالرموز) :	١	٪	:	*

(٢) البيانات الترتيبية (Ordinal) Data

تعرف البيانات الترتيبية بالبيانات المرتبة فى فئات أو حسب خصائصها عن طريق إعطاء القيم الأصلية للمتغيرات رتباً أو أرقاماً تدرجية أو تنازلية. فمثلاً عند تصنيف المناطق المختلفة فى الوجه البحرى - مصر - حسب كمية الأمطار التى تسقط عليها، نجد أن هناك مناطق تسقط عليها أمطاراً كميتها أقل من ٢٥ ملليمتر، ٢٥ - ٧٥، ٧٥ - ١٢٥، أكثر من ١٧٥ ملليمتر، وفى هذه الحالة تعطى المنطقة التى تسقط عليها أقل من ٢٥ ملليمتر الرقم (١) والمنطقة التى

تسقط عليها ٢٥ - ٧٥ ملليمتر الرقم (٢)، وهكذا إلى أن نصل إلى المنطقة التي تسقط عليها أكثر من ١٧٥ ملليمتر فيعطى لها الرقم (٤). وتتبع نفس الأسلوب في حالات أخرى كقياس الانحدارات وتقسيمها إلى انحدارات شديدة، انحدارات معتدلة، انحدارات طفيفة، وهكذا. ويستخدم هذا النوع من طرق قياس البيانات أيضاً مع الظواهر التي تختلف طبيعة متغيراتها، والتي يمكن أن توضع في شكل تتابع أو تسلسل تدريجي قائم على أساس العلاقة بينها. ويعد مقياس موس Mohs's scale لتعيين صلابة المعادن من أوضح الأمثلة للبيانات الترتيبية. وقد اختير لهذا المقياس عشرة معادن رتب ترتيباً تصاعدياً حسب درجة صلابتها النسبية. وتبدأ بالمعدن الأقل صلابة وهو التلك ودرجة صلابته النسبية (١) وتنتهي بالمعدن الأعظم صلابة وهو الماس ودرجة صلابته (١٠)، وعلى أساس هذا الترتيب في الصلابة يمكن تقدير صلابة المعادن الأخرى تقديراً نسبياً. وعلى العموم، إذ علم الترتيب (التدرج) الحقيقي للفئات التي تجمع على أساسها البيانات، فإنه يمكن إعطاء أى أرقام تدريجية لتمثل الفئات. فمثلاً تتكون الرواسب الشاطئية من مجموعة من العناصر يمكن ترتيبها حسب حجم الحبيبات التي تتكون منها وإعطائها أرقاماً تدريجية كما يلي:

اسم العنصر :	صلصال	غرين	رمل	حصى
أو :	٤	٣	٢	١
الحجم :	٨٥٦	٢٧٩	٥٢	١
أو :	١	٢	٣	٤

(٣) بيانات الفترة Interval Data :

تعد بيانات الفترة أكثر أنواع البيانات الإحصائية شيوعاً واستخداماً في أبحاث العلوم الأرضية ومن بينها الجغرافية. وبيانات الفترة تعكس القيم الأصلية للظواهر في شكل فئات لها أطوال (أو فترات) فيما بينها، كأعمار السكان في فئات السن المختلفة، وبالمثل كميات الإنتاج الزراعي والصناعي، مساحات المزارع، كميات الأمطار، ودرجات الحرارة، فكلها بيانات يمكن أن تقاس على أساس فئات

محدودة. ويتميز هذا النوع من البيانات بأن النسبة بين أى فترتين مستقلين من وحدات القياس ونقطة صفر القياس لأحد مقياس الظاهرة تكون مساوية لمثلتها على مقياس آخر لنفس الظاهرة. فمثلاً، على الرغم من إختلاف صفر المقياس المثوى عن صفر المقياس الفهرنهايتى لدرجة الحرارة، فإن نسبة أى فترتين على أحد المقياسين تساوى نفس النسبة على المقياس الآخر.

درجات مئوية :	صفر	٢٠	٥٠	١٠٠
درجات فهرنهايت :	٣٢	٦٨	١٢٢	٢١٢

$$\text{النسبة لفترتين (مئوية)} = \frac{٥٠ - ١٠٠}{٢٠ - \text{صفر}} = ٢,٥$$

$$\text{النسبة لفترتين (فهرنهايتية)} = \frac{١٢٢ - ٢١٢}{٣٢ - ٦٨} = ٢,٥$$

ومن المعروف أن هناك بعض الاختبارات الإحصائية التى لا تقبل إلا بيانات الفترة، بل أن معظم الأساليب الإحصائية مثل: تحليل التباين، معاملات الارتباط، تحليل الانحدار، تشترط أن تكون البيانات من نوع بيانات الفترة.

(٤) بيانات النسبة Ratio Data :

يمكن اعتبار بيانات النسبة من نوع بيانات الفترة التى تقاس على مقياس نسبى (Ratio Scale) والتى تكون فيها درجة صفر المقياس ذات قيمة أو درجة حقيقية (True Zero). وتعد قياسات كل من بيانات الكتلة Mass الأطوال، الأوزان، سرعة التيار النهري، زوايا ميل وخط مضرب الطبقات الصخرية، من أمثلة هذه البيانات. فمثلاً إذا كان أحد الكتب يزن رطلاً وكتاب آخر يزن رطلين، فإن نسبة وزنيهما تكون $\frac{١}{٢}$ ، وإذا استخدمنا المقياس المترى فإن وزنيهما سيكون ٤٥٣ جرام و ٩٠٦ جرام على الترتيب. وفى هذه الحالة ستظل النسبة بينهما ٠,٥ $(\frac{٤٥٣}{٩٠٦} = \frac{١}{٢})$. أما بالنسبة لبيانات الفترة التى لا يمكن

قياسها على أساس نسبي، فمن أمثلتها درجة الحرارة. فمثلاً إذا كانت درجة الحرارة عند وقت الظهيرة ليومين متتاليين هي ١٠ مئوية و ٢٠ مئوية، فإن النسبة بينهما هي $\frac{10}{20} = 0,5$ ولكن ستختلف هذه النسبة لنفس درجتى الحرارة بالمقياس الفهرنهايتى أى $\frac{50}{98} = 0,51$.

ومما تجدر الإشارة إليه أن طريقة القياس النسبى للبيانات تعتبر أكثر تنوعاً وقوة من أى طريقة قياس أخرى تجمع على أساسها البيانات الإحصائية، لأنها تحقق الحصول على أكبر كمية من المعلومات، كما أنها تسمح بتطبيق أكثر الأساليب الإحصائية دقة.

مما سبق يمكن القول أن بيانات الصفات المميزة Attributes للظواهر تقاس بمقياس البيانات الأسمية والترقية وتأخذ قيما غير متصلة أو وثابة. أما المتغيرات Variables فتقاس بمقياس بيانات الفترة والنسبة وتأخذ قيماً مستمرة (Yale and Kendall, 1953).

الدقة والأخطاء فى البيانات Precision and Errors

يشير معنى الدقة Precision فى الإحصاء إلى التكرار فى عملية القياس، والتي ينتج عنها قيمة تقديرية تمثل القيمة المتوسطة لعدد مرات قياس المتغير موضع الدراسة. وكلما كانت طريقة القياس مضبوطة (دقيقة) كلما كانت المفردات الناتجة عن عملية القياس قريبة من بعضها أو تقترب جميعها من قيمة متوسطة Average Value. ونظراً لأن تكاليف وغرض عملية القياس غالباً ما تتحكم فى مقدار الدقة المطلوبة فإنه ليس بالضرورة أن تكون القياسات الدقيقة لمتغير ما صحيحة أو مضبوطة Accurate. ويعزى السبب فى ذلك إلى أن القياس الصحيح أو المضبوط يتوقف على تقدير قرب القيمة الدقيقة من القيمة الحقيقية، أو بمعنى آخر عدم تحيز القياس، الظاهرة المقاسة. وفى الدراسات الجغرافية تواجهنا مشكلة عدم معرفة القيم الحقيقية للظاهرة. ولكن فى بعض الحالات يمكن افتراض قيمة معيارية تقارن بها بقية القيم أو القياسات. فعلى سبيل المثال يمكن قياس انحدار جزء من الأرض

طوله ٥٠٠ متراً بدقة كبيرة إلى أقرب ملليمتر، ومع ذلك إذا كانت أداة القياس (الشريط) غير مرقمة ترقيماً صحيحاً، أى نتج عنها خطأ أو تحيز، فإننا نحصل على قياس دقيق ولكنه غير صحيح. وتبعاً لذلك يتعرض العمل الإحصائي إلى أنواع كثيرة من الأخطاء Errors، أثناء تنفيذه، وسنشير هنا إلى نوعين رئيسيين من أنواع الأخطاء التى يتعرض لها قياس البيانات، والتى من شأنها التأثير على النتائج التى نحصل عليها من العينات وهما: أخطاء التحيز Bias Errors والأخطاء الاحتمالية Probability Errors.

وأخطاء التحيز هى الأخطاء الناجمة عن تدخل الباحث فى طريقة اختيار العينة. فالمعروف مثلاً أن العينة العشوائية تمثل بشكل كبير خصائص المجتمع الذى سحبت منه. فإذا اختيرت العينة بطريقة شخصية (أى غير عشوائية)، فإن ذلك يؤدي إلى زيادة الأخطاء المتوقعة. كذلك تنشأ هذه الأخطاء نتيجة لتحيز الباحث لوجهة نظر خاصة تجاه القرارات المتخذة. ويحدث خطأ التحيز عادة فى اتجاه واحد إما بالزيادة أو بالنقص. ويمكن أن تعزى أخطاء التحيز لعدة عوامل أهمها:

- ١- الاختيار المتعمد (غير العشوائى) للعينة.

- ٢- استبدال مفردات العينة بمفردات أخرى لعدم تمكن الباحث من الوصول لبعض المفردات الأساسية فى العينة.

- ٣- سوء التقدير وعدم توفر الدقة Precision، فقد لا يوفق الباحث فى التفرقة بين ما هو سبب أو نتيجة، أو عدم توفر الدقة فى حصر وحساب المتغيرات المحددة لطبيعة الظاهرة، ووضع فروض غير سليمة.

أما الأخطاء الاحتمالية فهى الأخطاء الناجمة عن احتمالات عدم تماثل النتائج التى نحصل عليها مع خصائص المجتمع. فحتى عندما تؤخذ العينة بالأسلوب العشوائى فإنه تظل هناك احتمالات أخطاء فى مدى تمثيل العينة لخصائص المجتمع الذى أخذت منه. ومن أهم هذه الأخطاء ما يطلق عليه إحصائياً «خطأ

الصدفة، أو «الخطأ العشوائي». وهو الخطأ الذى ينشأ فى عملية اختيار العينة. فإذا فرضنا أننا أخذنا من مجتمع الطلبة فى أحد قاعات الدراسة أربعة أفراد كانت أعمارهم كالآتى: ٢١، ١٩، ١٧، ٢٢ نجد أن متوسط عمر الفرد فى المجموعة المختارة هو $\frac{79}{4} = 19,75$ سنة. ونفرض أننا رمزنا له بالرمز ل. فإذا أخذنا من بين هؤلاء الأربعة ثلاثة أفراد عشوائياً، نجد أن هناك طرقاً عددها $4 \times 3 = \frac{4!}{1!3!}$ أى $\frac{24}{6} = 4$ طرق. فلا يمكن إذن سحب العينة إلا بإحدى الطرق الأربعة التى تكون المجموعات (١، ٢، ٣)، (١، ٢، ٤)، (١، ٣، ٤)، (٢، ٣، ٤). وإذا حسبنا متوسط عمر الفرد فى كل المجموعات الأربع الممكنة نجده على الترتيب ١٩، ٢٠، ٢٠، ٢٣. فإذا رمزنا إلى كل متوسط من هذه المتوسطات بالرمز ع، نجد أن ع تختلف عن ل إما بالزيادة أو بالنقص كالآتى، وعلى الترتيب: $-0,75 + 0,92 + 0,25 - 0,42$ ويلاحظ أن المجموع الجبرى لهذه الاختلافات يساوى صفراً، وهذا يعنى أن الاختلاف (ع - ل) يكون فى المتوسط لكل العينات الممكنة مساوياً للصفر.

ويطلق على اختلاف التقدير (ع) عن القيمة الحقيقية أو قيمة المجتمع بالخطأ العشوائى أو خطأ الصدفة لأن هذا الخطأ يرجع إلى عملية اختيار العينة كما ذكرنا. وعلى أية حال يتوقف الخطأ العشوائى على عاملين رئيسيين هما:

١- حجم العينة: فكلما كبرت العينة كلما قل الخطأ العشوائى، وزادت الثقة فى النتائج التى نحصل عليها.

٢- تباين مفردات المجتمع: فكلما زاد تباين مفردات المجتمع زاد احتمال حدوث الأخطاء العشوائية. وأياً كان الأمر، لو اتبع الأسلوب العشوائى فى اختيار العينة لأصبح من السهل والإمكان تقدير هذا النوع من الخطأ من العينة ذاتها. وسيتضح ذلك جلياً عند دراسة تقدير معالم المجتمع Population parameters فى الفصول القادمة.

الباب الأول

جمع البيانات وطرق إعدادها للتحليل الكمي

مقدمة

الفصل الأول: جمع البيانات

الفصل الثاني: تصنيف وجدولة البيانات

الفصل الثالث: العرض البياني للبيانات

الباب الأول

جمع البيانات وطرق إعدادها للتحليل الكمي

مقدمة:

يمر أسلوب أو خطة العمل الإحصائي أو ما يطلق عليه أحياناً «العمليات الإحصائية» بأربع مراحل رئيسية هي:

(١) مرحلة جمع البيانات Data Collection عن الظاهرة (الظواهر) موضع البحث من مصادرها المتنوعة سواء بواسطة الجهود الشخصية للمباحث، أو عن طريق البيانات المنشورة.

(٢) مرحلة جدولة وعرض البيانات Data Tabulation and Presentation بما تتضمنه من طرق تفريغ المعلومات بأساليب وأشكال (جداول، رسوم أو أشكال بيانية ... إلخ) تعكس بشكل واضح وبسيط خصائص الظاهرة موضع الدراسة.

(٣) مرحلة تحليل البيانات Data Analysis بما تشمله من متغيرات يتأثر بعضها البعض، وعلاقات متداخلة مع بعضها البعض. ويتم التحليل بواسطة المقاييس الإحصائية والأساليب الكمية المتنوعة.

(٤) مرحلة تفسير البيانات Data Intepretation، وهي أوج المراحل الثلاث السابقة، كما أنها تهدف إلى معرفة العوامل التي تتحكم في تنوع الظاهرة وتغير سلوكها.

وكل مرحلة من هذه المراحل تتضمن عدة طرق وأساليب في تنفيذها. وسوف نقوم بتوضيح وتفصيل المرحلتين: جمع البيانات، جدولة وعرض البيانات في فصول هذا الباب من الدراسة، أما بالنسبة للمرحلتين: تحليل البيانات وتفسير البيانات فسيتم شرحهما تبعاً في فصول الأبواب اللاحقة.

الفصل الأول جمع البيانات

Data Collection

الفصل الأول

جمع البيانات

Data Collection

تعتبر مرحلة جمع البيانات والمعلومات والحقائق عن المتغيرات والظواهر موضع الدراسة من أسس العمل الإحصائي التي لها أهمية خاصة لا يمكن إغفالها في أى دراسة علمية منظمة. وقبل الشروع فى عملية جمع البيانات يجب أن يلم الباحث بعدة خطوات هامة وضرورية تمليها عليه طبيعة الدراسة يمكن أن نوجزها فيما يلى:

- أ- تحديد المشكلة العلمية أو تعيين مجال الظاهرة المراد دراستها وبحثها.
 - ب- الإتفاق على وحدة القياس التى ستستعمل فى عملية جمع البيانات.
 - ج- تعيين المتغيرات التى ستناولها عملية القياس والحصر.
 - د- حصر المصادر التى يعتمد عليها فى الحصول على البيانات.
 - هـ- تحديد الأسلوب أو الطريقة التى تتبع فى جمع البيانات والمعلومات.
- وسوف نركز مناقشتنا فى هذا الفصل حول الإطار العام لكيفية جمع البيانات من مصادرها المختلفة وما يتصف به كل مصدر من مزايا الاستخدام ومثالب ومشاكل التطبيق. وتجدر الإشارة هنا إلى أنه كلما كانت طريقة جمع البيانات سليمة، وكلما توفرت معلومات دقيقة عن مجموعة المتغيرات أو الظاهرة موضع الدراسة، كلما أدى ذلك إلى رفع درجة الثقة فى النتائج المستخلصة من التحليل الإحصائي، وبالتالي التوصل إلى قرارات سليمة غير متحيزة.

مصادر جمع البيانات Sources of Data

هناك مصدران أساسيان لجمع البيانات: الأول، يستمد منه الباحث المعلومات اللازمة لبحثه من بيانات تم جمعها وتجهيزها ونشرها بواسطة أجهزة متخصصة أما المصدر الثانى فيعتمد فيه الباحث على نفسه فى جمع وإعداد وتجهيز البيانات. ويعرف المصدر الأول بالمصدر غير المباشر، بينما يطلق على المصدر الثانى «المصدر المباشر» أو مصدر الميدان.

أولاً: المصدر غير المباشر فى جمع البيانات:

تتصف البيانات التى نحصل عليها من هذا المصدر بأنها بيانات غير أولية، تم تبويبها وتصنيفها من قبل بواسطة شخص آخر (غير الباحث) أو هيئة حكومية، ومن أمثلتها البيانات التى تتضمنها الدوريات والنشرات والكتب والتقارير والبحوث التى تصدرها وتنشرها الجهات والهيئات الحكومية ومراكز البحوث العلمية. ويلجأ الباحث إلى هذا المصدر فى الحصول على البيانات التى يحتاج إليها بحثه فى حالة وجود صعوبات (من حيث الوقت والتكاليف) تعترض عملية جمع البيانات من مصادرها الأولية. وعلى الرغم من سهولة وسرعة الحصول على البيانات من هذا المصدر، إلا أنه يعاب عليه صعوبة تحديد درجة الدقة أو الثقة فى البيانات، وعدم التأكد من سلامة الإعداد والتجهيز الإحصائى لها. وللتغلب على كل ذلك على الباحث أن لا يعتمد فى الاعتماد على هذا المصدر فى حصوله على البيانات، وإذا كان مضطراً لذلك فيجب عليه الاعتماد على البيانات التى تصدرها أجهزة الإحصاء الرسمية فى الدولة، مثل الجهاز المركزى للتعبئة العامة والإحصاء بجمهورية مصر العربية.

ثانياً: المصدر المباشر فى جمع البيانات:

تتميز البيانات التى يتم الحصول عليها من هذا المصدر بأنها بيانات أولية يعتمد الباحث فى جمعها وتجهيزها للتحليل على نفسه. ويلجأ الباحث إلى هذا المصدر

فى حالة إذا ما كانت طبيعة الدراسة تملئ عليه الحصول على بيانات غير منشورة، أو نتائج بحوث سابقة تتعلق بموضوع البحث، كما فى دراسة العلاقة بين العمليات البحرية Marine Processes (الأمواج، التيارات ... إلخ) والظواهر التى تتأثر بها على ساحل منطقة ما فى وقت معين. ومن مزايا المصدر المباشر فى الحصول على المعلومات أن درجة الدقة وحدود الثقة فى البيانات يمكن تحديدها عند تحليل البيانات كمياً، وهى فى الغالب ما تكون مرتفعة مما يساعد بالتالى على استخلاص نتائج موثوق فيها بدرجة كبيرة. إلا أن أهم المشاكل التى تواجه الاعتماد على المصدر المباشر هو الحاجة إلى الوقت والتكلفة المادية اللازمين لإنجاز مهمة الحصول على المعلومات. ونتيجة لذلك فإن الباحث يجد نفسه مضطراً إلى بذل قصارى جهده فى جمع البيانات التى يحتاج إليها بالطريقة المباشرة فى وقت قصير وبأقل تكلفة مادية ممكنة.

وعند جمع البيانات من مصادرها المباشرة فإن الباحث يعتمد على أحد الأسلوبين: أسلوب الحصر (المسح) الشامل لجميع مفردات المجتمع الأصلي، فإذا لم يتيسر له ذلك فإنه يضطر إلى اختيار عينة، وهذا ما يطلق عليه أسلوب المعاينة (العينات). ولكل من الأسلوبين جوانبه الإيجابية والسلبية التى نوضحها فيما يلى:

أولاً: أسلوب الحصر (المسح) الشامل:

يعرف أسلوب الحصر الشامل أحياناً بأسلوب العد الكامل (أو التعداد Census) حيث أن معظم التعدادات تتم من خلاله، مثل التعداد السكاني Population Census والتعداد الزراعى أو التجارى أو الصناعى التى يعتمد عليها فى استخراج بعض المقاييس والمؤشرات الاحصائية، والتى تكون أساساً فى عملية التخطيط القومى أو وضع إطار عام للأبعاد الفعلية لإمكانية الدولة فى مواجهة الأزمات الاقتصادية أو الاجتماعية وغيرهما. والأساس فى عملية جمع البيانات عن طريق الحصر الشامل هو إدخال كل مفردات المجتمع الاحصائى، دون استبعاد أى مفردة، فى البحث والاستقصاء. فمثلاً عند دراسة العمالة الصناعية فى محافظة ما يقوم الباحث بعمل

حصر شامل لجميع العمال حسب نوع كل صناعة، وكذلك عند دراسة التركيب المحصول للأحواض الزراعية في أحد مراكز محافظة ما فإن الباحث يقوم بعمل حصر شامل لأنواع المحاصيل والمساحة التي تشغلها داخل كل حوض من الأحواض الزراعية. وبناء على ذلك فإن هذا الأسلوب يطبق عند دراسة المجتمعات الاحصائية مجهولة المعالم والتي تتطلب جمع بيانات شاملة عن كل مفردة من مفردات المجتمع حتى يمكن تحديد خصائصه ومعالمه بكل دقة وبدرجة عالية من الثقة.

ولأسلوب الحصر الشامل بعض المثالب والمشاكل عند استخدامه في جمع البيانات فهو لا يصلح للأبحاث التي يقترن استخلاص النتائج منها بوقت محدد، أو بمعنى آخر أن هذا الأسلوب لا يتناسب مع الأبحاث التي يكون فيها لعنصرى الوقت والتكاليف المالية أهمية خاصة وأثر كبير على استخلاص النتائج. وعلاوة على ذلك يتعرض تنفيذ أسلوب المسح الشامل في جمع البيانات لكثير من الأخطاء التي من أهمها خطأ تحيز الباحث، سواء كان تحيز متعمد أو غير متعمد، الذى ينجم عن أخذ كل مفردات المجتمع في الدراسة حيث وجود احتمالات الخطأ في العد أو احتمالات تجاهل بعض المفردات مما يؤثر على دقة النتائج. وللتخلص من خطأ هذا الأسلوب يمكن تقسيم المجتمع إلى مجموعات متجانسة لها خصائص متشابهة ومميزات مترادفة، ثم يجرى البحث وعملية الحصر على كل قسم على حدة مع مراعاة التنسيق في الدراسة بين كل الأقسام. وأخيراً فإن هذا الأسلوب يتطلب فى اجرائه توفر جهاز فنى احصائى كبير واعتمادات مالية ضخمة ووقت متسع، مما يفسر أن معظم الدراسات والأبحاث التى تعتمد على هذا الأسلوب فى انجازها لا يقوم بها سوى أجهزة الإحصاء فى الدول مثل الجهاز المركزى للتعبئة العامة والإحصاء بجمهورية مصر.

ثانياً: أسلوب المعاينة (العينات) Sampling

سبق أن عرفنا أن دراسة المجتمعات الاحصائية تعتمد أساساً على أخذ كل

مفردات المجتمع للتعرف على خصائص ومعالم هذا المجتمع . وبصفة عامة فإن معالم أى مجتمع (وهى مقادير ثابتة للمجتمع الواحد ولكنها تتغير من مجتمع إلى آخر) هى التى تعطى لهذا المجتمع صفاته دون غيره. ونظراً لوجود صعوبات كثيرة تحول دون دراسة جميع مفردات المجتمع بواسطة أسلوب الحصر الشامل، فإننا نجري دراستنا على جزء صغير من هذا المجتمع أو ما يسمى بالعينة Sample، وذلك اختصاراً للوقت وتوفيراً للجهد والنفقات. واتباع دراسة العينات أو أسلوب المعاينة يرفع من مستوى العمل البحثي ويجعله أكثر دقة، وذلك لأن دراسة عدد قليل من المفردات أو الحالات يتيح للباحث فرصة جمع معلومات دقيقة وكثيرة عن كل مفردة أو حالة. وعلى العموم فإنه إذا ما وجدنا أنه من الضروري إجراء معاينة فإن رائدنا الأساسى يكون دائماً هو الحصول على عينة تعطى نتائجاً ذات دقة معينة بأقل تكاليف ممكنة، أو التى تعطى أعلى دقة بتكاليف محدودة.

ويفضل استخدام أسلوب المعاينة عند دراسة خصائص ومعالم المجتمعات اللانهائية مثل الوحدات الإنتاجية لإنتاج بعض الآلات، كما يفضل كذلك فى الأبحاث العلمية التى تتطلب تصور عام أو رأى عام حول قضية أو مشكلة يرى دراستها فى مجالات العلوم الطبيعية أو الاجتماعية. وفى كل من هذه الحالات يجب أن تكون العينات ممثلة تماماً للمجتمع ولا تخضع للاختيار الشخصى، وذلك حتى يمكن الحصول، بواسطة تطبيق الأساليب الكمية والمقاييس الاحصائية، على نتائج يمكن تعميمها على المجتمع الأصلي المراد تحديد معالمه بدرجة عالية من الدقة والثقة. ويجدر الإشارة هنا إلى أنه عند دراسة العينات فإن المقاييس التى تحسب من توزيع العينة المختارة (مثل الوسط الحسابى والانحراف المعياري - سيأتى ذكرهما فيما بعد بالتفصيل) يسمى كل منها «احصائية» وقيمة كل «احصائية» تختلف من عينة إلى أخرى. وللتفرقة بين المقاييس التى تحسبها من العينة وتلك التى نحصل عليها من دراسة جميع مفردات المجتمع بطريقة الحصر الشامل، تسمى

الأولى «بالاحصائيات» Sample statistics، بينما تعرف الثانية «بالمعالم» Population Parameters.

ويتوقف نجاح استخدام وتطبيق أسلوب المعاينة على عدة أمور هامة هي: تقدير حجم العينة، كيفية اختيار مفردات العينة من المجتمع، وتحديد نوع العينة. وفيما يلي مناقشة تفصيلية لكل منها على حدة.

(١) تقدير حجم العينة:

تتفق آراء كثير من الاحصائيين على أن حجم عينة البحث يتوقف على مجموعة من العوامل تنحصر في: الغرض من البحث، حجم المجتمع الأصلي، مدى تباين الظواهر المختلفة في قطاعات المجتمع. درجة الدقة المطلوبة في البحث، البيانات المتاحة التي يمكن استخدامها في تعميم النتائج، والامكانيات المادية. ونظراً لعدم وجود اتفاق بين الباحثين على وضع حد معين على أساس علمي - أو إحصائي - يحدد الحجم المناسب أو الأمثل للعينة لكي تمثل المجتمع الذي تسجب منه تمثيلاً جيداً، فإن تقدير حجم العينة - على مستوى معظم الدراسات والبحوث - يعتبر واحدة من المشكلات الخاصة بأسلوب المعاينة وتطبيق الأساليب الإحصائية. وفي مجال العمل الإحصائي يوجد اتجاهان عند تقدير حجم العينة.

الاتجاه الأول: يعتمد على الخبرة السابقة للباحث في هذا المجال، حيث أظهرت خلاصة الخبرات والتجارب أن حجم عينة في حدود ١٠٪ إلى ١٥٪ من حجم المجتمع الأصلي يبدو ملائماً في معظم الدراسات والبحوث. ويتميز هذا الاتجاه في تقدير حجم العينة بسهولة كما أنه يفيد بعض الباحثين قليلي الخبرة في مجال العمل الإحصائي.

الاتجاه الثاني: يرتبط أساساً بنظرية الاحتمال Theory of probability، مما يتطلب من الباحث الإلمام بقدر وافر من المعلومات الإحصائية والرياضية حتى يستطيع استخدام الأساليب الإحصائية في تقدير الحجم الأمثل للعينة. ويعتمد هذا

الاتجاه على تحديد العوامل (المتغيرات) التي يتوقف عليها حجم العينة واعتبارها دلائل رئيسية أو مؤشرات أساسية لهذا الغرض، وهو أمر يغفله الاتجاه الأول تماماً. كما يعتمد هذا الاتجاه على توفر بعض المعلومات عن حجم ومعالج المجتمع الأصلي عن طريق العينات التجريبية أو الاسترشادية - Experimental or Pilot Sample. وتتمثل أهم العوامل والمتغيرات الرئيسية المحددة لحجم العينة في: نسبة الخطأ المسموح به (أو درجة الدقة أو الثقة)، معامل التشتت (أو الإنحراف المعياري) بين مفردات العينة أو المجتمع إن أمكن، والاختلاف النسبي بين المتوسط الحسابي للعينة ومتوسط المجتمع. وتوضع هذه المتغيرات في شكل صيغة رياضية تختلف باختلاف حجم العينة الاسترشادية، كما تترجم على هيئة معادلة خاصة في حالة إذا كان حجم المجتمع الأصلي الذي ستحسب منه العينة معلوماً.

فإذا ما تصورنا أن أحد الجغرافيين بصدد تقدير حجم عينة من مجتمع كبير غير محدود المفردات فإنه يقوم بسحب عينة استرشادية من هذا المجتمع وحساب بعض المقاييس الاحصائية منها لتقدير بعض خصائص أو معالم المجتمع، والتي عن طريقها يمكن تقدير حجم العينة المطلوب. فإذا كان حجم العينة الاسترشادية ٣٠ مفردة أو أكثر فإن أهم العوامل المحددة لحجم العينة المطلوب تتمثل في:

أ- الانحراف المعياري بين مفردات العينة أو الخطأ المتوقع لمتوسط قيم مفردات العينة، ومنه يمكن تقدير الانحراف المعياري للمجتمع، أو ما يعرف «بأحسن تقدير Best Estimate للانحراف المعياري بين مفردات المجتمع، ويرمز له بالرمز $\hat{\sigma}$ ويحسب على أساس:

$$\hat{\sigma} = \sigma \times \sqrt{\frac{n}{n-1}} \quad \text{أو} \quad \sqrt{\frac{\sum (s - \bar{s})^2}{n-1}} \quad \text{أو} \quad \sqrt{\frac{\sum s^2 - \frac{(\sum s)^2}{n}}{n-1}}$$

حيث σ هي الانحراف المعياري للعينة، s هي قيمة مفردة من مفردات العينة، \bar{s} هي المتوسط الحسابي للعينة، n هي الحجم الفعلي للعينة.

ب- خطأ المعاينة أو الخطأ المعياري Sampling Error or Standard Error للمتوسط بين مفردات العينة أو مفردات المجتمع إن أمكن. وهو عبارة عن الدقة المطلوبة للتقدير الإحصائي من بيانات العينة، إذ أن تحديد حجم العينة يعتمد على الدرجة التي عندها يتجه متوسط العينة إلى الاختلاف والتباين عن متوسط المجتمع^(١). ويرمز للخطأ المعياري بالرمز (خ.م) ويحسب على أساس:

$$\sqrt{\frac{\hat{\sigma}^2}{n}} \times \sqrt{f-1} = \text{أو} \sqrt{\frac{\hat{\sigma}^2}{n}} \times \sqrt{f-1} = \text{(خ.م)}$$

$$\sqrt{\frac{\hat{\sigma}^2}{n} \times (f-1)} = \text{أو}$$

حيث ف تمثل نسبة حجم العينة إلى حجم المجتمع الأصلي $\left(\frac{n}{N}\right)$ وتسمى هذه النسبة «نسبة المعاينة» Sampling Fraction أو معامل التصحيح لقيمة الخطأ المعياري للمجتمع الأصلي الذي يجب أن يكون أقل من خطأ المعاينة للمتوسط. وكلما كان زاد حجم العينة واقترب من حجم المجتمع الأصلي كلما اقتربت ف من الوحدة (الواحد الصحيح) وأصبحت قيمة معامل التصحيح صفراً، وبالتالي فإن قيمة الخطأ المعياري تصبح صفراً أيضاً.

ج- القيمة المعيارية لاحتفال وقوع خطأ مسموح به ويرمز لها بالرمز (ز). ويمكن تحديد هذه القيمة من جدول التوزيع الاحتمالي الطبيعي إذا كان مستوى الثقة Confidence Level الذي تعمم به النتائج على المجتمع معلوماً.

وإذا أخذنا في الاعتبار المتغيرات الثلاثة السابقة فإن حجم العينة يمكن أن يتحدد في ضوء تحديد الفارق الممكن التسامح فيه بين نتيجة العينة، وما هو كائن

(١) مستنار مناقشة الخطأ المعياري Standard Error مناقشة تفصيلية عند دراسة أساليب المقارنة فيما

فعلاً في المجتمع Tolerance (أى الخطأ المعياري)، ومستوى الثقة التي تعمم بها النتائج على المجتمع. ويمكن وضع هذا التصور لحجم العينة بحسب أن:

$$\begin{aligned} \frac{\hat{\sigma}}{\sqrt{n}} &= (م.خ) \\ \sqrt{n} \times م.خ &= \hat{\sigma} \\ \frac{\hat{\sigma}}{م.خ} &= \sqrt{n} \\ \left(\frac{\hat{\sigma}}{م.خ} \right)^2 &= n \quad \dots \dots \dots (1-1) \end{aligned}$$

فإذا استطعنا أن نضع تقديراً مبدئياً للخطأ المعياري (م.خ) الذي نرغب أن ننتهي إليه، وإذا استطعنا أيضاً سحب عينة استرشادية كبيرة (من الثابت احصائياً أنه إذا بلغ حجم العينة ٣٠ مفردة أو أكثر فإنه يمكن أن يعطى حدوداً مرتفعة من الثقة لتقدير متوسط المجتمع وانحرافه المعياري من متوسط العينة وانحرافها المعياري، ويرجع ذلك إلى أنه كلما كان حجم العينة كبيراً كلما أدى ذلك إلى تكوين توزيع طبيعي للعينة يتركز حول المتوسط الحقيقي للمجتمع حتى لو كان توزيع المجتمع غير طبيعي)، فإنه باستطاعتنا تقدير الحجم الأمثل بالاستعانة بالمعادلة السابقة (١-١).

مثال ١: على أساس عينة استرشادية مكونة من ١٠٠٠٠ فدان قدر أن متوسط إنتاج الفدان من القمح على مستوى الجمهورية هو ٥ أرباب. وكان أحسن تقدير للانحراف المعياري ($\hat{\sigma}$) لإنتاج القمح على المستوى القومي ٢ أرباب، وأن الخطأ المعياري للمتوسط هو ٠,٢ أرباب. فلو افترضنا أننا نريد تقدير المتوسط القرمي لإنتاج الفدان من القمح لأقرب ١/٤ أرباب عند مستوى احتمالي ٦٨٢، (أى بمستوى الثقة ٦٨٪) فإن أقل حجم مطلوب للعينة في هذه الحالة يكون:

$$\text{حجم العينة (ن)} = \left(\frac{\hat{E}}{م.خ} \right)^2$$

$$= \left(\frac{2}{0.25} \right)^2 = 64 = ٦٤ \text{ فداناً}$$

وعلى ذلك فإن أى عينة مكونة من ٦٤ فدان تكون كافية لإعطاء تقدير للمتوسط القومى (أى متوسط المجتمع) بدقة $\pm 1/4$ أردب وبمستوى ثقة ٦٨,٢٪. ولو افترضنا، مرة أخرى، أن درجة الدقة المطلوبة لتقدير المتوسط القومى لإنتاج الفدان من القمح هى نفس الدقة السابقة (أى إلى أقرب $1/4$ أردب) ولكن عند المستوى الاحتمالى ٩٥,٤٪ (أى بمستوى الثقة ٩٥,٤٪) الذى تكون عنده حدود الثقة عبارة عن \pm خطأين معيارين للمتوسط (أى $\pm 2 \times م.خ$). ومعنى ذلك أن $2 \times$ قيمة الخطأ المعيارى للمتوسط لابد أن تساوى الدقة المطلوبة لحساب المتوسط العام وهى ٠,٢٥ أردب. وبعبارة أخرى فإن الخطأ المعيارى للمتوسط، لدرجة ثقة ٩٥,٤٪، يساوى $\frac{2}{0.25}$ أى ٠,١٢٥ أردب. وتطبيق المعادلة (١-١) فإن حجم العينة يكون فى هذه الحالة.

$$\text{حجم العينة (ن)} = \left(\frac{\hat{E}}{م.خ} \right)^2$$

$$= \left(\frac{2}{0.125} \right)^2 = 256 = ٢٥٦ \text{ فداناً}$$

وبناء على ذلك فإننا لكى نحصل على تقدير للمتوسط القومى لإنتاج الفدان من القمح لأقرب $1/4$ أردب بمستوى ثقة ٩٥,٤٪ فإن حجم العينة اللازم لابد أن يكون ٢٥٦ فداناً.

وهناك صيغة أو معادلة أخرى لتحديد الحجم الأمثل للعينة تأخذ فى اعتبارها المتغيرات السابق ذكرها والتى تحسب من عينة استرشادية يبلغ حجمها ٣٠ مفرداً أو أكثر، وهذه الصيغة هى:

$$\text{حجم العينة} = \left(\frac{\text{الانحراف المعياري للعينة} \times \text{القيمة المعيارية لاحتساب خطأ مسموح به بدرجة معينة}}{\text{الدقة المطلوبة للتقدير الإحصائي أو الخطأ المعياري}} \right)^2$$

أى أن:

$$n = \left[\frac{z \times \sigma}{\epsilon} \right]^2 \quad \dots \quad \dots \quad \dots \quad (2 - 1)$$

مثال ٢: إذا كان الانحراف المعياري لعينة استرشادية مكونة من ٣٠ عاملاً لدراسة مستوى المعيشة لمجتمع عمالي هو ١٠ جنيهاً شهرياً وأن الخطأ المعياري المسموح به لتقدير المتوسط العام للدخل الشهري هو ٢,٥ جنيهاً وذلك بمستوى ثقة ٩٥,٤٪ فإن الحجم الأمثل للعينة الذى يحقق الدقة المطلوبة يمكن تقديره بعد تحديد قيمة (z) المعيارية من جدول التوزيع المعتدل المناظرة لمستوى الثقة ٩٥,٤٪ وهى فى هذه الحالة تساوى ٢ تقريباً. وعلى ذلك فإن:

$$\text{حجم العينة (ن)} = \left[\frac{2 \times 10}{2.5} \right]^2 = 2(8) = 64 \text{ عاملاً}$$

وبما أن حجم العينة الاسترشادية ٣٠ عاملاً فإننا نحتاج إلى عدد ٣٤ عاملاً آخر ليصبح الحجم الفعلى للعينة ٦٤ عاملاً والذى منه يمكن تقدير المتوسط العام لدخل المجتمع العمالي قيد البحث بالدقة المطلوبة (أو الخطأ المعياري للمتوسط) المشار إليها).

ويمكن أيضاً تحديد حجم العينة التى تحقق الدقة المطلوبة أو الخطأ المسموح به فى حساب المتوسط العام للمجتمع من إحصائية عينة تجريبية صغيرة يقل عدد مفرداتها عن ٣٠ مفردة. وفى هذه الحالة تؤخذ العوامل السابقة المحددة لحجم العينة فى الاعتبار مع اختلاف واحد وهو أن القيمة المعيارية المناظرة لاحتساب وقوع

خطأ مسموح به بدرجة معينة في جدول التوزيع المعتدل يجب أن تستبدل بقيمة معيارية أخرى من جدول توزيع «ستيو دنت - ت» Student-t Distribution مناظرة لعدد من المفردات يقل عن مفردات العينة التجريبية بمفردة واحدة وعند مستوى الدلالة أو الثقة المطلوب.

وبناء على ذلك فإن:

$$\text{حجم العينة} = \frac{\text{مربع (الانحراف المعياري للعينة} \times \text{قيمة ت المعيارية)}}{\text{مربع (الدقة المطلوبة أو الخطأ المعياري)}} \quad \text{أى أن:}$$

$$n = \frac{(x \times z)^2}{(x_m)^2} \quad \dots \quad \dots \quad \dots \quad (3-1)$$

مثال ٣: فى دراسة اجتماعية عن مستوى المعيشة لعمال إحدى الصناعات التى يبلغ عدد مصانعها ٩٢ مصنعاً، يراد معرفة حجم عينة من المصانع يمكن منه تقدير متوسط الدخل السنوى لجميع العمال وذلك فى حدود ٥٠ جنيهاً زيادة أو نقص عن متوسط دخل عمال مصانع العينة بمستوى ثقة ٩٥% (أى بمستوى دلالة ٥%)، فإننا فى هذه الحالة نأخذ عينة تجريبية مكونة من دخول عمال ٢٥ مصنعاً ونحسب منها متوسط الدخل والانحراف المعياري ولنفترض أنه كان ٣٣٢ جنيهاً و ١٦١,٥ جنيهاً على الترتيب. وبما أن عدد مفردات العينة التجريبية أقل من ٣٠ مفردة، فلا بد إذن من تعيين قيمة ت من جدول توزيع «استيودنت - ت» المناظرة لعدد مفردات يقل عن مفردات العينة بمفردة واحدة (أى ٢٥ - ١) وعند نسبة الخطأ المسموح بها ٥%. وباستخدام هذه المؤشرات فإن قيمة «ت» تساوى ٢,٠٦. وبذلك فإننا يمكن أن نطبق المعادلة (٣-١) لنحصل على حجم العينة المطلوب كما يلى:

$$44,2 = \frac{2(2,06 \times 161,5)}{2(90)} = n$$

وبناء على ذلك فإن حجم عينة مكونة من عمال ٤٥ مصنعاً تكون كافية لإعطاء صورة صادقة عن الدخل السنوى لعمال جميع المصانع. ومن المعادلة السابقة (١-٣) يمكن أن تستنتج أنه كلما كبرت قيمة الانحراف المعياري للعينة التجريبية كلما زاد حجم العينة المطلوبة، والعكس يحدث مع الدقة المطلوبة أو الخطأ المعياري لحساب المتوسط العام، أى أنه كلما قلت (انخفضت) هذه الدقة أو زادت قيمة الخطأ المعياري كلما قل حجم العينة. ففي المثال السابق إذا انخفضت الدقة فى حساب المتوسط العام لتصل إلى ١٠٠ جنيهاً زيادة أو نقصان من متوسط العينة، فإن حجم العينة المطلوب سيكون ١٢ مصنعاً فقط، وهو حجم كاف أيضاً لإعطاء صورة عامة عن متوسط الدخل السنوى لعمال جميع المصانع بالدقة المطلوبة.

كما يمكن تحديد حجم العينة عن طريق تحديد النسبة المئوية لوجود الظاهرة موضع الدراسة فى العينة التجريبية وتعيين مستوى الثقة التى تعمم بها النتائج على المجتمع، وتحديد الفارق الممكن التسامح فيه بين نتيجة العينة وما هو كائن فعلاً فى المجتمع، وهذا يتطلب تحديد الخطأ المعياري للعينة على النحو التالي:

$$\sqrt{\frac{p \times q}{n}} = \text{الخطأ المعياري (خ.م.)}$$

حيث أ هى النسبة المئوية لوجود الظاهرة، ب هى النسبة المئوية لعدم وجود الظاهرة أى ١٠٠ مطروحاً منها نسبة وجود الظاهرة، ن هى حجم العينة. ويمكن نقل هذه المعادلة على النحو التالي:

$$n = \frac{p \times q}{2(\text{خ.م.})^2} \dots \dots \dots (1 - \epsilon)$$

وفى كل الحالات إذا استطعنا أن نضع تقديراً مبدئياً لدقة القياس أو الخطأ المعيارى الذى نرغب أن ننتهى إليه، وإذا استطعنا كذلك تقدير نسبة وجود الظاهرة فى الحالات المدروسة، فإنه باستطاعتنا تقدير الحجم الأمثل للعينة بالمعادلة السابقة (١ - ٤).

مثال ٤ : من عينة استرشادية لدراسة مدى تأثير البرامج التلفزيونية على ثقافة سكان أحد الأقسام الإدارية بمحافظة ما وجد أن النسبة المئوية لحائزى الأجهزة التلفزيونية هي ٨٦٪ بدقة تصل إلى $\pm ١٠\%$ ، والمطلوب تحديد الحجم الأمثل للعينة التى يمكن عن طريقها دراسة هذا التأثير بمستوى ثقة ٩٥,٤٪. فيما أن مستوى الثقة التى تعمم بها النتائج على المجتمع هو ٩٥,٤٪ فإن حدود هذه الثقة Confidence limits عبارة عن \pm خطأين معيارين (أى $\pm ٣ - \text{خ.م.}$) وقيمتيهما لابد أن تساوى ١٠٪. ومعنى ذلك أن الخطأ المعيارى فى هذه الحالة هو ٥٪ (أى $\frac{10}{2}$)، وأن حجم العينة المطلوب حسب المعادلة (١ - ٤) هو:

$$\begin{aligned} \text{حجم العينة (ن)} &= \frac{p \times q}{\frac{1}{2}(\text{خ.م.})^2} \\ &= \frac{14 \times 86}{\frac{1}{2}(5)^2} = \\ &= \frac{1204}{25} = 48,2 \end{aligned}$$

وعلى ذلك فإن الحجم الأمثل للعينة والذى نأمل أن يحقق الدقة المطلوبة فى هذا المثال هو ٤٩ حائزاً لأجهزة التلفزيون فى المنطقة موضع الدراسة.

وتستخدم صيغة أخرى لتحديد الحجم المناسب للعينة اعتماداً على خصائص بيانات عينة استرشادية يمكن منها تعيين النسبة المئوية لوجود الظاهرة موضع الدراسة بالإضافة إلى تقدير الخطأ المعيارى وتحديد مستوى الثقة التى تعمم بها النتائج. هذه الصيغة هي:

$$ن = أ \times ب \times \left[\frac{ز}{م \times خ} \right]^2 \dots \dots (٥ - ١)$$

حيث ز هى القيمة المعيارية لاحتمال وقوع خطأ مسموح به عند مستوى احتمالى معين:

مثال ٥: من عينة تجريبية تتكون من ٣٠ ناخباً وجد أن ١٢ ناخباً سيقومون باعطاء أصواتهم لمرشح الحزب (أ)، فإذا أريد تقدير نسبة الناخبين الذين سيدلوا بأصواتهم لانتخاب مرشح هذا الحزب من جملة الناخبين بدقة (أو خطأ معيارى) $\pm ١\%$ وبمستوى ثقة ٩٥,٤% فإن حجم العينة المطلوب لتحقيق ذلك يتحدد على أساس:

$$أ- نسبة الناخبين فى العينة = \frac{١٠٠ \times ١٢}{٣٠} = ٤٠\%$$

ب- القيمة المعيارية (ز) لمستوى الثقة ٩٥,٤% = ٢ تقريباً.

ج- الخطأ المعيارى للتقدير = ١%

وعلى ذلك فإن:

$$حجم العينة (ن) = ٦٠ \times ٤٠ \times \left(\frac{٢}{١} \right)^2$$

$$= ٢٤٠٠ \times ٤ = ٩٦٠٠ مفردة أو ناخباً.$$

وتجدر الإشارة هنا إلى أن معظم استطلاعات الرأى Opinion Polls فى النواحي السياسية تعتمد على عينات يصل حجم أى منها إلى ٢٠٠٠ مفردة تقريباً، حتى يكون الخطأ المعيارى المتسامح فيه بين نتيجة العينة وما هو كائن فعلاً فى المجتمع $\pm ٢\%$ أو أكثر قليلاً، وذلك قبل أن يدخل فى الاعتبار عوامل أخرى مثل خطأ التحيز أو تغيير الرأى فى آخر دقيقة تجاه موضوع الاستطلاع (مثل ترشيح عضو أحد الأحزاب الأخرى) من جانب نسبة من المبحوثين أو الناخبين.

كما يمكن تحديد حجم العينة عن طريق معرفة حجم المجتمع الأصلى فقط،

وذلك عن طريق تحديد مجموعة من العوامل أو المحددات الرئيسية التى يمكن أن نجعلها فيما يلى:

أ- حجم المجتمع الأصلي الذى ستسحب منه العينة ويرمز له بالرمز N .

ب- معامل الاختلاف بين مفردات العينة ويرمز له بالرمز (m) ويحسب على أساس:

$$(m) = \frac{\text{الانحراف المعياري}}{\text{المتوسط الحسابي}} \times 100$$

ج- مربع متوسط معامل الاختلاف المتوسط بين مفردات العينة ويرمز له بالرمز (m^2) ويحسب على أساس:

$$(m^2) = \left[\frac{m}{n} \times (1 - f) \right]^2$$

حيث n هى حجم العينة، f هى نسبة المعاينة Sampling Fraction أى نسبة حجم العينة إلى حجم المجتمع الأصلي.

د- الفارق النسبى بين المتوسط الحسابى للعينة ومتوسط المجتمع ويرمز له بالرمز (q) ، ويمكن حسابه كما يلى:

الفارق النسبى = الجذر التربيعى لمربع متوسط معامل الاختلاف بين مفردات العينة \times القيمة المعيارية للدقة المطلوبة بدرجة معينة

أى أن :

$$q = (m^2 \times z)$$

وبناء على العلاقات الرياضية المتبادلة بين هذه العوامل الأربعة وحجم العينة فإننا يمكن أن نحدد حجم العينة المطلوب فى ضوء المعادلة الآتية:

حجم العينة =

$$\frac{\text{حجم المجتمع الأصلي} \times \text{مربع القيمة المعيارية} \times \text{مربع معامل الاختلاف}}{\text{حجم المجتمع الأصلي} \times \text{مربع الفارق النسبى} + \text{مربع القيمة المعيارية} \times \text{مربع معامل الاختلاف}}$$

أى أن:

$$n = \frac{2(m) \times (z) \times 0.5}{2(m) \times 2(z) + 2(q) \times 0.5} \dots (1 - 6)$$

مثال ٦: يريد أحد الباحثين تحديد حجم عينة من مجتمع محتوى على ٢٠٠٠ مفردة وذلك فى ضوء الافتراضات الآتية التى يرى أنها ضرورية لتطبيق الطرق الإحصائية واستخلاص النتائج التى على أساسها تتخذ القرارات اللازمة.

أ- معامل الاختلاف بين مفردات العينة (م) فى حدود ٢٠٪.

ب- متوسط معامل الاختلاف بين مفردات العينة (م س) فى حدود ٢٪.

ج- نسبة الخطأ المسموح به لارتفاع عن ٥٪ أى أن تعميم النتائج بثقة قدرها ٩٥٪.

د- بما أن القيمة المعيارية المناظرة لاحتمال وقوع الخطأ المسموح به وهو ٥٪ فى جدول التوزيع المعتدل (الطبيعى) تساوى ٢ تقريباً فإن:

$$\text{الفارق النسبى (ق)} = 0.2, \quad 2 \times 0.4 = 0.8$$

وعلى ذلك فإن حجم العينة المطلوب يمكن تحديده كما يلى:

$$\text{حجم العينة (ن)} = \frac{2(0.2) \times 2(2) \times 2000}{2(0.2) \times 2(2) + 2(0.4) \times 2000}$$

$$95.25 = \frac{320}{3.36} =$$

وعلى ذلك فإن حجم العينة المناسب الذى يحقق افتراضات الباحث هو ٩٦ مفردة تقريباً من مجتمع يحتوى على ٢٠٠٠ مفردة.

وهناك طريقة أخرى لتحديد حجم العينة يمكن استخدامها إذا كان حجم المجتمع الأصلى معروفاً، وذلك بعد تحديد الحجم التقريبى للعينة الذى يتطلب:

أ- تحديد الدقة المطلوبة أو الخطأ الذى يمكن التسامح بين نتيجة العينة وما هو كائن فعلاً فى المجتمع.

- ب- تحديد مستوى الثقة التى تعمم بها النتائج على المجتمع.
- ح- اختيار النسبة المئوية لوجود الظواهر قيد البحث (ح) التى تحقق أكبر رقم إذا ما ضربت فى النسبة المئوية المكملة (١٠٠ - ح).
- وتتخذ معادلة هذه الطريقة الشكل الآتى:

$$n = \frac{[\frac{(100 - \%) \times \text{ح}}{(\text{م.ح})^2}] \times (z)^2}{(1 - \%) \dots}$$

حيث ح هى نسبة وجود الظواهر قيد البحث وتمثل ٥٠٪، ن_١ هى الحجم التقريبى للعينة.

وبعد الحصول على الحجم التقريبى للعينة، يتعين تحديد الحجم الفعلى لها فى ضوء حجم المجتمع الأصلى (٠.٣) وذلك باستخدام معادلة التصحيح وهى.

$$\text{الحجم الفعلى للعينة (ن)} = \frac{\frac{n}{1 - \frac{n}{N}} + 1}{0.3} \dots (1 - 1)$$

مثال ٧: لنفرض أن أحد الباحثين يريد تحديد حجم عينة من مجتمع يحتوى على ٤٠٠٠ مفردة بناء على بعض الافتراضات التى رآها ضرورية فى هذا الصدد. هذه الافتراضات هى:

- أ- نسبة الخطأ المسموح به (الدقة) فى حدود $\pm ٥\%$.
- ب- مستوى الثقة التى تعمم بها النتائج لا تقل عن ٩٥٪.
- ح- نسبة وجود الظاهرة موضع البحث فى العينة ٥٠٪ ونسبة عدم وجودها ٥٠٪ أيضاً.

وباستخدام هذه الافتراضات، والتى يمكن أن تتحقق من تحديد حجم مناسب للعينة، نجد أنه عند مستوى الثقة ٩٥٪ تكون القيمة المعيارية المناظرة فى جدول التوزيع المعتدل (الطبيعى) هى ٢ تقريباً.

. وعلى ذلك فإن حجم العينة المطلوب يتحقق بتطبيق المعادلتين (١-٧)،
(١-٨) كما يلي:

$$\text{الحجم التقريبي للعينة} = \frac{٥٠ \times ٥١}{٢(٥)} \times ٢(٢)$$

$$٤٠٠ = \frac{٤ \times ٢٥٠٠}{٢٥} = \text{مفردة}$$

$$\text{الحجم الفعلي للعينة} = \frac{\frac{٤٠٠}{١ - \frac{٤٠٠}{٤٠٠٠}} + ١}{\frac{٤٠٠}{٤٠٠٠}}$$

$$٣٦٣,٩٤ = \frac{٤٠٠}{١,٠٩٩} = \text{مفردة}$$

وعلى ذلك فإن حجم العينة المناسب والذي يمكن أن يحقق افتراضات الباحث هو ٣٦٤ مفردة من مجتمع يحتوى على ٤٠٠٠ مفردة أى بنسبة ٩.١٪ تقريباً من حجم المجتمع الأصلي.

أما إذا تجمعت لدينا بيانات خاصة عن معالم المجتمع الأصلي المتوسط الحسابي والانحراف المعياري الذي ستسحب منه العينة دون أن نتمكن من معرفة حجمه، فإننا نستخدم طريقة المنحنى المعتدل (الطبيعي) للحصول على الحجم المناسب أو الأمثل للعينة. ويوضح ذلك المثال التطبيقي الآتي:

مثال ٨: إذا كان متوسط الإنتاج القومي للقمح في عام ما هو ٧,٢ أردب للفدان والانحراف المعياري لهذا المتوسط هو ١,٥ أردب، فما هو حجم العينة بالفدان التي ينبغي اختيارها من محافظة ما بشرط ألا يكون خطأ الصدفة أكثر من ٥.٠٪ وأن يكون متوسط الإنتاج في العينة ٧,٠ أردب؟

نظراً لأننا افترضنا أن نسبة خطأ الصدفة هي ٥.٠٪، فإنه يمكن إيجاد القيمة المعيارية (ز) والتي تعرف بالمتغير المعياري Standard Variable من الجداول

الإحصائية الخاصة بتحديد المساحة تحت المنحنى المعتدل (الطبيعي) ^(١) المناظرة
لقيمة خطأ الصدفة. ثم تطبق المعادلة الآتية:

$$\frac{\text{الفرق بين متوسطي المجتمع والعينة} \times \sqrt{\text{حجم العينة}}}{\text{الانحراف المعياري لمتوسط المجتمع}} = \text{المتغير المعياري}$$

أى أن

$$(z) = \frac{\bar{m} - \bar{s} \times \sqrt{n}}{e} \quad \dots \quad \dots \quad (1 - 9)$$

وعلى ذلك فإن:

$$\frac{\sqrt{n} \times 7,0 - 7,2}{1,5} = 1,64$$

$$2,46 = \sqrt{n}$$

$$12,3 = \frac{2,46}{,2} = \sqrt{n}$$

$$n = 51,29 \text{ فداناً}$$

بمعنى أنه ينبغي أن تسحب عينة حجمها ٥٢ فداناً من المحافظة قيد البحث
لكي تحقق الدقة أو خطأ الصدفة والمتوسط المطلوب للعينة.

من كل مما سبق يمكن القول أن حجم العينة الذى نحصل عليه بإحدى
المعادلات السابقة لا يعتبر ملزماً لأن الافتراضات التى تقوم عليها هذه المعادلات غير
ملزمة لأى دراسة، وكل ذلك ما هو إلا مجرد علامات تحدد أسلوب العمل فى
هذا المجال فى حدود أقل خطأ ممكن وبطريقة موضوعية غير متحيزة.

(١) انظر جدول تحديد المساحة تحت المنحنى المعتدل (الطبيعي) ضمن ملاحق هذا الكتاب.

(٢) اختيار مفردات العينة:

بعد أن تعرفنا على الطرق المختلفة التي تحدد الحجم المناسب للعينة التي سيجرى عليها البحث والاستقصاء، فإننا الآن بصدد التعرف على طريقة (عملية) اختيار مفردات هذه العينة من بين مفردات المجتمع الأصلي، أو ما يعرف بأسلوب سحب العينة من المجتمع. وعملية اختيار مفردات العينة، كواحدة من المشكلات الخاصة بأسلوب المعاينة، تتوقف أساساً على حجم المجتمع الأصلي. فإذا كان حجم المجتمع صغيراً أي مشتملاً على عدد محدود Finite من المفردات، فإن المشكلة لا تكون مشكلة اختيار العينة من بين مفردات المجتمع، بل تكون مشكلة الحصول على كاف من المفردات لغرض البحث. فمثلاً إذا أراد الباحث أن يجرى دراسة على كبار المزارعين بإحدى القرى، كنموذج لنفس الفئة في القطر كله، فقد يحدد هذه الفئة بأنها تشتمل على كل من يمتلك ١٠٠ فداناً أو أكثر من الأراضي الزراعية في القرية. وفي هذه الحالة قد يكون عدد هؤلاء الملاك قليلاً لدرجة أن العينة تستنفذهم جميعاً. كما تكون عملية الاختيار من المجتمع الأصلي عملية مشترطة بشروط تحدد المفردات (عدد الملاك) التي تتكون منها العينة المطلوبة. وبالطبع كلما كثرت الشروط اللازمة للعينة كلما صعب الحصول عليها وكلما قل عدد المفردات التي يتم الاختيار من بينها. أما إذا كان حجم المجتمع الأصلي كبيراً جداً أي مشتملاً على عدد غير محدود من المفردات المستوفية لجميع الشروط اللازمة في العينة فإنه من اللازم إجراء عملية اختيار مفردات العينة إما بواسطة الاختيار غير العشوائي (المعاينة العمدية) أو بواسطة الاختيار العشوائي. وقبل أن نوضح كيفية الاختيار في كل من الطريقتين، فإنه يجدر بنا أن نذكر الشروط التي ينبغي توافرها في العينة حتى نستعيز بها عن المجتمع الأصلي الكبير.

سبق أن قلنا أنه يجب أن تكون العينة ممثلة للمجتمع أو بمعنى آخر يجب مراعاة أن تمثل العينة جميع مفردات المجتمع، وألا تكون متحيزة biased لجزء أو

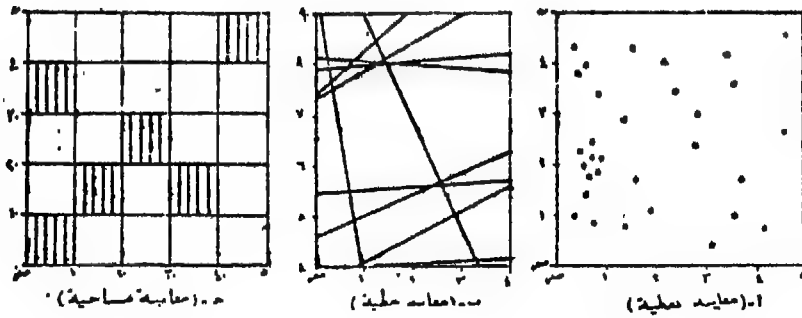
أجزاء من المجتمع الأصلي لأنه يتوقف على العينة المنتقاة كل قياس أو نتيجة يخرج بها الباحث. فمثلاً إذا أردنا سحب عينة لتقدير متوسط الدخل من الإنتاج الزراعي لملاك الأراضي الزراعية على مستوى أحد المراكز، فإذا أخذت عينة من فئة الملاك الذين يملكون ١٠٠ فداناً أو أكثر فإن العينة تكون غير ممثلة لمجتمع الملاك حيث أن هذه الفئة تمثل نسبة صغيراً جداً من جميع الملاك، وبالتالي لا بد أن تحتوى العينة على ملاك من جميع فئات الملكية. وبناء على ذلك يجب أن تتوفر في العينة الممثلة Representative Sample مجموعة من الشروط يمكن تلخيصها في شرطين أساسيين هما:

١- أن تكون مفردات العينة ممثلة للمجتمع الذى يجرى عليه البحث تمثيلاً صحيحاً، وليست ممثلة لمجتمع آخر. بمعنى أنه إذا تكررت نفس النتائج على عينات أخرى من نفس المجتمع كانت العينة التى يجرى عليها البحث عينة ممثلة للمجتمع الأصلي أصدق تمثيل. وبذلك يمكن أن تكون خصائص مفردات العينة (احصائيات العينة) متقاربة أو متشابهة مع خصائص المجتمع (معالم المجتمع) الذى تنتمى إليه.

٢- ألا تكون المفردات المختارة ممثلة لجزء (قطاع) من أجزاء المجتمع الأصلي بل يجب أن تمثل جميع أجزاء المجتمع، وهذا يتطلب تكوين إطار المعاينة الذى تؤخذ منه العينة.

وإطار المعاينة Sampling Frame هو المصدر الذى تؤخذ منه العينة. أو بعبارة أخرى هو حصر شامل (القائمة أو الدليل) لجميع مفردات أو وحدات المجتمع الأصلي المراد دراسته. مثال ذلك قائمة بأسماء العمال فى أحد المصانع، أو مختلف أنواع الرواسب التى توجد على الشاطئ، أو مواقع المحلات العمرانية الريفية على خريطة إحدى الدول. وعند اختيار العينة من المجتمع المحدودة يقسم المجتمع الأصلي للظاهرة قيد البحث إلى عدة أقسام تسمى وحدات المعاينة Sampling Units (شخص، أسرة، قرية) ويكون إطار العينة حينئذ عبارة عن القائمة أو مجموعة

القوائم التى تتضمن الوحدات التى يتألف منها المجتمع. ويشترط فى إطار المعاينة أن يكون شاملاً لجميع مفردات المجتمع التى يمكن الوصول إليها بسهولة وذلك حتى يكون اختيار العينة سليماً. كما يشترط أن يكون إطار المعاينة متجديداً حتى تعطى المفردات أو الوحدات التى تستجد على الإطار القديم نفس الفرصة فى الظهور فى العينة. وكما ذكرنا فإنه فى المجتمعات غير المحدودة Infinit يستحيل إجراء حصر شامل لكل مفردات المجتمع فى الوقت المتاح للدراسة، ويكتفى فى هذه الحالة بدراسة عينة بدون تكوين إطار للمعاينة. ويلاحظ على إطار المعاينة فى مجال الدراسة الجغرافية أنه إما أن يكون إطاراً مكانياً Spatial أو غير مكانياً Non - Spatial. ويقصد بإطار المعاينة المكانى بأنه الإطار الذى يكون فيه المكان Location هو الوحدة الرئيسية، كما أنه الأساس فى اختيار العينات التى تمثل التغيرات (الاختلافات) المكانية Locational Variability التى يتميز بها مجتمع الأماكن لمنطقة ما تمثيلاً صحيحاً. فمثلاً إذا كنا بصدد معاينة خريطة بهدف تحديد مواقع الأماكن التى يشغلها نوعاً معيناً من النشاط البشرى على هذه الخريطة، فإننا يجب أن نتأكد من تمثيل كل أجزاء الخريطة تمثيلاً صحيحاً. ويتم ذلك باختيار أحد أنواع المعاينة الآتية (شكل رقم ١ - ١).



شكل رقم (١)

أنواع إطار المعاينة الجغرافية

أ- المعاينة النقطية Point - Sampling أى معاينة نقط مواقع الأماكن على خريطة المنطقة.

ب- المعاينة الخطية Line - Sampling أى بأخذ عينة من قطاعات عرضية مختلفة من الخريطة.

ج- المعاينة المساحية Area - Sampling أى بأخذ عينة تمثل مساحة مجموعة من المربعات التى تغطى مساحة خريطة المنطقة قيد البحث.

وعلى ذلك يكون إطار المعاينة عبارة عن جميع مفردات المجتمع لكل شكل من أشكال المعاينة الثلاثة، أو بمعنى آخر جميع النقط المحددة لمواقع الأماكن، أو القطاعات العرضية، أو المربعات التى تغطى مساحة الخريطة. وعلى الرغم من أن طبيعة عمل الجغرافى عند جمعه للبيانات ترتبط بإطار المعاينة المكانى، إلا أنه فى بعض الأحيان ولظروف خاصة نجده يهتم بتحديد إطار معاينة غير مكانى ليلتزم دراسته. فمثلاً إذا كان يصدد اختيار عينة من أسر أحد الأقسام الإدارية فى مدينة ما وذلك لتقدير متوسط الدخل، فإنه يتحتم عليه اختيار عينة من إطار (أو قائمة) يحتوى على جميع أسر هذا القسم الإدارى بالمدينة. ولا يجوز له فى هذه الحالة أن يختار العينة من دليل التليفون مثلاً إذا أنه من المعروف أن مثل هذا الدليل لا يتضمن جميع أسر القسم الإدارى قيد البحث.

الاختيار غير العشوائى (العمدى)؛

يلجأ الباحث أحياناً إلى اختيار مفردات عينة بطريقة غير عشوائية أو متعمدة، فمثلاً قد يختار الباحث عينة يرى أنها تمثل المجتمع بالنسبة إلى صفة أو خاصية ما دون غيرها. وبعبارة أخرى يكون الأساس فى الاختيار هو الباحث الذى يحدد بنفسه المفردات الداخلة فى العينة متحيزاً لتفكيره ومتعمداً فى تحديد المفردات. فمثلاً إذا أراد باحث دراسة مستوى المعيشة فى الريف المصرى فقد يعتقد أن قرية معينة فى نظرة تمثل مستوى المعيشة فى كل الريف المصرى، وفى هذه الحالة إذا اختار هذه

القرية كعينة أو أساس للدراسة فإن وجهة نظر الباحث تكون غير صائبة ولن يحالفها التوفيق في استخلاص النتائج لأنه باختياره لهذه القرية يكون معتمداً أو متحيزاً في الاختيار لغرض الدراسة ويمكن القول أن الباحث في تعمله في اختيار مفردات العينة إنما يتخلى عن فكرة العشوائية في اختيار مفردات العينة من مفردات المجتمع. على أننا يجب ألا نقلل من أهمية الاختيار العمدى فربما يكون هو أفضل الطرق عند الاختيار في حالة إذا ما كان المطلوب اختيار عينة صغيرة لمجتمع كبير. فإذا كان المطلوب اختيار قرية واحدة لتمثيل القطر المصرى كله فإنه يمكن اعتبار الاختيار العمدى هو أفضل الطرق رغم ما فيه من تحيز، وذلك لأن اختيار قرية واحدة بطريقة عشوائية قد يؤدي إلى خطأ كبير.

الاختيار العشوائى:

على الرغم من سهولة اختيار أو سحب عينة غير عشوائية من المجتمع كله إلا أن ذلك له أضراره البالغة على دقة النتائج تبعاً لوجود تحيز من قبل الباحث في اختيار مفردات العينة ولعدم توفر عنصر العشوائية في الاختيار. ولضمان الحصول على المعاينة غير المتحيزة التى تعطينا تقديرات واستنتاجات يمكن تعميمها على جميع مفردات المجتمع الأصلى بأعلى دقة لتكاليف محدودة، لا بد أن نتخير العينة بحيث تكون ممثلة تمثيلاً صادقاً للمجتمع ويكون ذلك على أساس نظرية الاحتمالات، أى بواسطة سحب وحدات بالتتابع لكل منها احتمال معروف في الاختيار في السحبات المختلفة. ولكى تكون مفردات العينة ممثلة لكل مفردات المجتمع الأصلى بأقل أخطاء ممكنة، فلا بد من استخدام أسلوب العشوائية غير المتحيز في الاختيار.

والأساس فى أسلوب الاختيار العشوائى للعينة هو أن جميع مفردات المجتمع موضع الدراسة لها نفس الفرصة فى الاختيار (Alder and Rossler, 1964) وهذا يعنى عدماً لإهتمام ببعض المفردات دون الأخرى وإتاحة الفرص المتكافئة أمام كل مفردة لتكون ضمن العينة. والمعنى الرياضى فى الاختيار العشوائى هو أن يكون

احتمال ظهور أى مفردة من مفردات المجتمع الأصلي فى العينة متساوى، ويساوى

$\frac{1}{\text{حجم المجتمع}}$ ، ويكون احتمال الحصول على مفردة من مفردات المجتمع فى العينة

يساوى $\frac{\text{حجم العينة}}{\text{حجم المجتمع}}$ [أى $\frac{n}{N}$]. وبذلك فإن الشرط الاحصائى الأساسى

لاختيار مفردات العينة من بين مفردات المجتمع هو عدم التحيز فى الاختيار حتى نضمن - إلى درجة ما - تمثيل العينة للمجتمع الذى نريد معاينته تمثيلاً صادقاً مع تقليل الأخطاء الأخرى للمعاينة.

وتتم عملية اختيار مفردات العينة بالأسلوب العشوائى بإحدى الطرق الآتية:

أ- طريقة السحب العشوائى (القرعة).

عند اتباع هذه الطريقة تعطى مفردات المجتمع الأصلي أرقاماً مسلسلّة تكتب على بطاقات متشابهة، وبعد أن تخلط خلطاً جيداً - يكفى لا ضاعة أى أثر للترتيب المتعمد - يسحب عشوائياً منها عدد عدد من البطاقات يساوى عدد مفردات العينة المطلوبة. وتلائم هذه الطريقة سحب العينات من المجتمعات الصغيرة الحجم حيث أنها لا تحتاج إلى مجهود كبير أو وقت طويل فى عملية ترقيم مفردات المجتمع الأصلي وسحب العينة منه.

ب- طريقة الجداول العشوائية:

يصعب اتباع طريقة السحب العشوائى فى الاختيار فى حالة المجتمعات الكبيرة الحجم حيث تحتاج عملية الترقيم إلى مجهود كبير كما أنها تكون مضيّعة للوقت. ولذلك يفضل الرجوع إلى جداول خاصة تعرف باسم جداول الأرقام العشوائية Tables of Random Numbers. وهى جداول أرقامها مختارة بطريقة عشوائية (أى أنها أرقام لا ترتبط ببعضها بأى أسلوب رياضى، فهى لا تكون بينها متتالية عددية أو هندسية) وموضوعة فى شكل أعمدة تتناسب مع حجم مجتمع أحصائى يتكون من أى عدد من المفردات كما يبدو من الجدول التالى (جدول رقم ١ - ١).

جدول رقم (١-١) جدول الأرقام العشوائية

٥٢	٩٢	٥٥	٥١	١٠	٨٦	١٠	٠٢	٦١	٣٨	٦٦	٥٩	١٧	٢٣	٢٨	٤٢	١٧	٢٠
٤٩	٩٤	٦٠	٩٤	٦٣	٤٨	٥٤	١١	٧٠	٥٣	٣٣	١٠	٠٤	٠٣	٤٩	٠٤	٤٩	٧٤
٦٤	٤٨	١٨	٣٧	٧١	٧٨	٨٨	٤٠	٦٥	٢٩	٤٢	٢٣	٦٧	٣٨	٣١	٤٩	٧٠	٩٤
٣٧	٣٨	٣٧	٠١	٠٢	٥٤	١٢	١٥	٥٤	٣٢	٥٢	٣٢	٨٤	٦٩	١٥	٧٨	١٥	٢٢
٥٣	١٢	٣٦	٤٩	٥١	٥٨	٥٧	٥٠	٨٧	٩١	٥٥	٣٠	٣٠	٢٧	١٨	١٢	٢٩	٩٣
٥٦	٨٥	٨٧	٥١	٨٣	٠٣	٨٥	٦٩	٩٥	٥٢	٤٥	٩٩	١٤	٣٦	٩٧	٧٧	٠٤	٤٥
٥١	٠٨	٢٦	٥٩	٧٣	٦٤	٧٧	٠٦	٤٩	٤٨	٦٠	٩٤	٣٩	٨٩	٤٩	٩٩	٩١	٤٤
٢٤	٢٦	٣٧	٦٣	٩٢	٣٠	٨٤	٢٧	٦٥	٨٩	٥٩	٤٧	٩٦	١٩	٠٢	٩١	٢٣	١٦
٣٤	٩٩	٨٣	٨٨	٤٧	٦١	٠٤	٥٥	٥١	٧٠	٤٢	٨٢	٦٥	٦٥	٠٤	٦٥	٥٠	٠٤
٩٢	٠٨	٧٨	١٧	٣٣	٨٨	٧٧	٢٢	٧١	٢٤	٢٦	٦٦	٦١	٠٣	٧٢	١٧	٧٠	٣٢
٠٨	٣٩	٩٠	٥١	٣٤	٩٢	٨١	٢٠	٤١	٠٦	٣٩	٨١	٩٥	٤٢	٠٧	٥٩	٦٤	٠٣
٤٧	٢٧	٠٣	٤٩	٦٨	٦٧	٠٧	١٩	٨٣	٣١	٤٨	٩٣	٨٦	٦٧	٩٠	٠٠	٤٩	٦٣
٧٦	٣٣	٨٦	٠٦	٤٠	٣٣	٠٧	٥١	٨٨	٤٨	٠٣	١٤	٣٦	٩٨	٨٦	٩٥	٠٠	٦١
٧٩	٠١	٨٠	٥٢	٠٣	٢٢	٤٥	٦٠	٩٦	٠٩	٠٤	٢١	٧٤	٢٨	٤٩	٩٠	٠٣	٨٩
٧٩	٨٥	٢٤	٥٥	٢٩	١٤	٢٧	٨٩	٧١	٠٦	٠٧	٦٠	٤٠	٥٢	٨٥	٣٣	٧٢	٠١
١٤	٩٣	٧٠	٤٤	٢٦	٣٣	١٧	٩١	٥٣	٦٠	٢٢	٣٣	٣٤	٣٤	٧٩	٤٩	٥٦	٣٧
٣٥	٢٦	٣٤	٦٦	٦٦	٣٥	٢٢	٢٠	٢٨	٢٤	٢٥	٣٥	١٥	١٠	٤٨	٧٤	٠٥	٤٩
٦٦	٠٣	٦٩	٩٢	٥٨	٥٩	٢٨	٠١	٢٣	٤٢	٩٤	٣٣	٢٦	٩٧	٢١	٣٧	٧٤	٤٩
٨٥	٩٧	٠٧	٥٦	٦٣	٦٢	٦٥	٩٧	١٢	٠٨	٨٥	١٩	٠٨	٨٨	٤٣	٢٢	٢٦	٣٠
٣٦	٩٧	٧٥	٩٣	٥٥	٥٤	١٨	٢٢	٧٩	٠٨	٩٣	٧٦	٣٩	٤٣	٩٦	٧٧	٨٧	٤٨
٦٧	٠٤	٢١	٢٢	٨٥	٣٠	٢٠	٠٥	٧	٢٧	١١	٠٠	٧٣	٧١	٤٦	٨٧	٧٢	٠٨
٩٩	٢٠	٣٢	١٩	٧٥	٣٢	٢٢	٤٦	٧١	٦٤	٤٢	٣١	٢٧	١٧	٦٢	٩٨	٩٧	٩٥
١٠	٧٢	٢٠	٦٩	٧٤	٣٦	٣٢	٢٤	١٢	٤٦	٥٥	٤٦	٤٠	٧٠	٣١	٥٧	٩٩	٣٧
٢٣	٩٦	٤٨	٠٠	٢٨	٥٤	٤٤	٢٣	٧١	٨٨	١٨	٧٥	٣٣	٨٥	٣٧	١٨	٧٩	٠٥
١١	٢٦	٦٥	٣٦	٤٠	٥١	٣٩	٤١	٠١	٢٩	٢٢	٩١	٧٩	٠٠	٤٢	٦٣	٨٥	٥٥

19	08	87	72	79	00	22	29	80	2	72	22	27	78	20	97	28	77
87	07	07	89	70	20	82	72	22	87	82	01	09	22	78	92	87	80
7	70	27	18	11	27	20	08	12	72	22	78	88	00	09	80	10	20
19	20	20	77	27	22	22	87	89	27	8	77	80	90	28	89	00	92
09	12	27	07	10	92	02	29	19	72	72	22	20	22	77	77	72	11
71	02	20	22	02	27	20	78	72	92	07	28	00	02	02	27	00	72
79	70	98	20	21	20	27	88	70	10	08	70	21	78	22	12	92	00
80	17	19	27	97	28	80	70	09	22	00	20	12	22	97	27	98	77
70	08	28	72	77	07	72	79	90	70	91	90	77	70	82	22	91	77
22	80	82	08	22	90	20	29	29	21	20	19	07	02	18	12	08	22
20	27	20	00	22	98	02	02	90	00	00	72	20	71	17	20	29	02
21	82	77	89	27	20	08	20	92	97	27	79	20	78	99	92	98	72
22	92	79	00	21	22	88	79	20	02	07	79	88	01	20	02	27	00
22	07	09	29	72	28	20	19	22	28	82	97	79	22	89	71	27	29
90	77	77	87	97	20	00	27	02	72	88	72	29	78	22	12	70	19
28	22	87	00	28	28	00	12	71	72	2	81	09	87	77	27	17	72
70	22	07	28	89	11	88	01	90	07	07	07	97	98	27	97	22	18
01	27	70	22	88	77	81	77	02	10	71	97	09	71	87	70	08	70
02	22	78	72	92	20	10	22	70	21	70	80	12	91	00	21	90	79
11	22	77	70	77	72	20	20	22	92	09	17	00	09	10	00	22	07
01	20	28	19	12	00	19	22	22	20	27	90	01	01	22	12	08	90
20	07	97	00	20	70	80	12	02	21	22	82	21	02	72	82	02	
72	09	22	22	21	29	82	02	22	20	70	77	00	02	10	27	17	98
22	09	20	89	17	70	02	72	00	20	72	20	20	82	22	11	91	08
00	90	72	11	21	22	27	97	02	80	19	77	87	82	77	27	21	27

وتعتمد طريقة استخدام جداول الأرقام العشوائية على حجم العينة المطلوب حبها من المجتمع الأصلي. فعند استخدام هذه الجداول تعطى أرقام مسلسل مفردات المجتمع الذى نريد معاينته، ثم يختار اختياراً عشوائياً بداية تؤخذ من عندها الأرقام التى تدخل ضمن مفردات العينة مع استبعاد الأرقام المتكررة أو الأرقام التى تزيد عن حجم المجتمع. ولتحقيق ذلك يلزم أخذ أعمدة تحتوى على عدد من الخانات تساوى عدد أرقام حجم المجتمع، ولشرح ذلك نقول: لو كانت هناك بيانات عن مجتمع يتكون من ١٠٠ مفردة فبإمكاننا اختيار الأرقام بالطريقة العشوائية من العمودين الأول والثانى من الجدول رقم (١-١). فإذا كنا نريد اختيار عينة مكونة من ١٠ مفردات من ١٠٠ مفردة، فإننا سنرى أن أرقام المفردات المطلوبة، من الجدول السابق، هى ٢٠، ٧٤، ٤٩، ٢٢، ٩٣، ٤٥، ٤٤، ١٦، ٤، ٣٢، أما إذا كان حجم المجتمع كبيراً كأن يتكون من ١٠٠٠٠ مفردة، ففي هذه الحالة تستعمل الأعمدة الأربعة فى الجدول المذكور. فإذا أردنا سحب عينة مكونة من ١٠٠ مفردة من هذا المجتمع، فإن أرقام مفردات العينة المختارة من الجدول بهذه الطريقة تكون هى على التوالى ١٧٢٠ ثم تليها ٤٩٧٤. وبعدها ٧٠٩٤ وهكذا حتى يصل عدد المفردات ١٠٠. وهى عدد مفردات العينة المطلوبة. وإذا كنا بصدد معاينة ٢٠٠٠ مفردة فقط فإننا نختار من الجدول السابق أرقام مكونة من أربعة حدود لمفردات العينة المطلوبة وهى ١٠٠ مفردة.

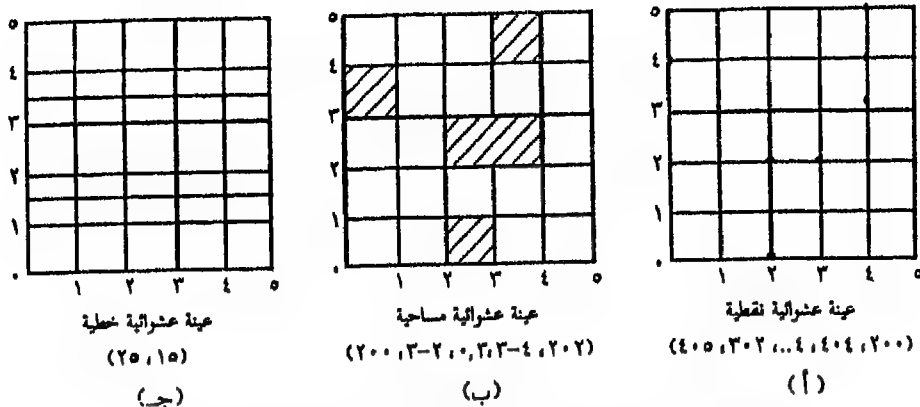
وتتم عملية الاختيار بإحدى الطريقتين: أما أن تختار الأرقام العشوائية من العمودية الأول والثانى لمفردات العينة بقبول أى رقم يقع بين ١ و ٢٠٠٠ ورفض أى رقم آخر يقع بين ٢٠٠٠ و ٩٩٩٩، ونستمر فى عملية القبول والرفض حتى يصل عدد المفردات ١٠٠ مفردة. ويعاب على هذه العملية أنها تستنفذ وقتاً طويلاً وذلك بسبب ارتفاع معدل رفض الأرقام التى تزيد عن الحجم الكلى للمفردات الذى نريد سحب العينة منه. ولهذا السبب تستبدل طريقة قبول ورفض الأرقام العشوائية بطريقة أخرى يمكن بها اختيار المفردات المطلوبة وذلك على أساس قبول

جميع الأرقام التي تزيد عن رقم ٢٠٠٠ واعتبارها تكرارات للأرقام بين ١ و ٢٠٠٠، بمعنى أن الأرقام العشوائية بين ٢٠٠١ و ٤٠٠٠ و ٤٠٠١ و ٦٠٠٠ و ٦٠٠١ و ٨٠٠٠ و ٨٠٠١ و ١٠٠٠٠ يمكن اعتبارها أرقاماً إضافية (جديدة) للأرقام من ١ إلى ٢٠٠٠. وتمتاز هذه الطريقة بأن فيها اختصار كبير للوقت في اختيار أرقام المطلوبة من جميع الأرقام المحصورة بين ١ و ١٠٠٠٠.

وفي بعض الحالات تكون البيانات المتاحة عن مجتمع الظاهرة قيد البحث في شكل مجموعات ويراد أخذ عينة عشوائية من المجموعات ككل وليس من كل مجموعة على حدة. فمثلاً إذا كانت لدينا بيانات عن أعداد السكان لعدد كبير من المراكز العمرانية التي تتكون من عدة مجموعات: قرية صغيرة، قرية كبيرة، مدن صغيرة، مدن كبيرة (عواصم المحافظات)، فإنه بإمكاننا أخذ عينة واحدة من كل هذه المجموعات باستخدام الطريقة العشوائية في الاختيار. ويتم ذلك بأن ترتب أعداد كل المجموعات في قائمة حتى نحصل على المجموع الكلي لأعداد جميع المراكز العمرانية المطلوب معاينتها. فمثلاً إذا كان عدد المراكز العمرانية للمجموعات الأربع فهو ١٣٦٠، ٦٧، ٣٢، ٧ على الترتيب، فإنها تأخذ ترقيماً من ١ إلى ١٣٦٠، ومن ١٣٧ إلى ٢٠٤، ومن ٢٠٥ إلى ٢٣٧، ثم من ٢٣٨ إلى ٢٤٥ على الترتيب أيضاً. وبالتالي يتكون المجتمع الاحصائي للمجموعات الأربع من ٢٤٥ مركزاً عمرانياً، والذي يمكن منه سحب أية عينة بالطريقة العشوائية السابق شرحها.

وبالإضافة إلى الاختيار العشوائي لمفردات العينة من بين مفردات المجتمع التي تحتويها قائمة بأعداد هذه المفردات، فإنه يمكن أيضاً تطبيق نفس أسلوب الاختيار على البيانات الخاصة بالتوزيعات المكانية Spatial (Areal) Distribution التي تعد حجر الزاوية في الدراسات الجغرافية. ويتطلب إجراء معاينة التوزيعات المكانية بطريقة الاختيار من الجداول العشوائية أن يرسم على خريطة المنطقة قيد البحث شبكة من المربعات، في حالة إذا لم يوجد على الخريطة شبكة خطوط الطول ودوائر العرض، ويعطى لخطوط شبكة المربعات (الطولية والعرضية) التي أنشأناها أرقاماً أو إحداثيات

من الغرب إلى الشرق ومن الجنوب إلى الشمال بحيث يكون رقم الصفر (بداية التقسيم أو الترقيم) في الركن الجنوبي الغربي (نقطة الأصل) للخريطة الذى تتزايد منه الاحداثيات شرقاً وشمالاً. ويمكن وضع الأرقام بحيث تدل على خطوط الاحداثيات نفسها أو تشير إلى المسافة بين كل خط إحداثى وآخر يليه، ويتوقف ذلك على نوع التوزيعات المكانية المطلوب معاينتها. فمثلاً عند الدراسة الجغرافية لبعض مظاهر النشاط الزراعى فى منطقة ما، كدراسة الدورات الزراعية والتي يكون الهدف منها معرفة التأثير الذى يفرضه استخدام دورة زراعية معينة على الإنتاج الزراعى (مع افتراض ثبات تأثير العوامل الأخرى)، فإننا فى هذه الحالة نختار مواقع الأماكن الزراعية (أو القرى وزمامها الزراعى) كإطار للمعاينة لأنه من المعروف أن تلك الأماكن تمثل المصدر الذى يستمد منه البيانات الخاصة لمثل هذه الدراسة. وبما أننا نريد اختيار مجموعة من القرى على خريطة كمية للدراسة فإننا نكون بصدد تطبيق أسلوب المعاينة النقطية Point - Sampling والذى يمكن اعتباره أفضل الطرق لهذا الغرض. ولتحديد مواقع القرى المطلوبة على خريطة لاتتعدى فيها خطوط الإحداثيات عن ١٠٠ خط فى كل اتجاه على حدة، نقوم بإختيار العمودين الأول والثانى أو أى عمودين آخرين من جدول الأرقام العشوائية (جدول رقم ١ - ١)، ونمثل برقمى العمود الأول المسافة على الإحداثى الشرقى (المحور الأفقى العشوائى) من نقطة الأصل، وبرقمى العمود الثانى المسافة على الإحداثى الشمالى (المحور الرأسى العشوائى) من نقطة الأصل أيضاً. أو بعبارة أخرى تدل الأرقام العشوائية المختارة على قيم أو أرقام خطوط الإحداثيات لكل محور على حدة. وعلى أساس الإحداثى (الشرقى والشمالى) العشوائى المستخرج من الجدول يرسم عمودان أحدهما على المحور الأفقى العشوائى والآخر على المحور الرأسى العشوائى حيث تحدد نقطة تلاقيهما موقع القرية التى تدخل ضمن مفردات العينة المطلوبة (شكل رقم ١ - ١٢).



شكل رقم (١-٢): أنواع الاختيار العشوائى للتوزيعات المكانية

أما إذا كنا نريد معاينة مظاهر استخدام الأرض Land Use فى منطقة ماء، فإننا نكون بصدد تطبيق أسلوب المعاينة المساحية Area - Sampling والذي هو أفضل الأساليب لمثل هذه المعاينة. ويتطلب إجراء المعاينة المساحية على خريطة إما أن ترقيم المسافات الإحداثيات (شكل رقم ١ - ٢ ب) بأرقام عشوائية تستخرج من الجدول وتحدد المساحات المطلوبة للعينة داخل المربعات بطريقة الإحداثيات السابق ذكرها، وإما أن نبقى على ترقيم خطوط الإحداثيات كما هى الحال فى طريقة المعاينة النقطية (شكل رقم ١ - ٢ ب) وتختار المربعات (أو المساحات) المطلوبة بطريقة الإحداثيات أيضاً.

ومن البيانات الجغرافية ما يختص بالتوزيعات المكانية الخطية (مثل خطوط أو شبكات المواصلات بأنواعها المختلفة، والأنهار وروافدها، أو أى ظاهرة جغرافية تتخذ شكل الامتداد الخطى) والتي يمكن معاينتها بالمعاينة الخطية Line - Sampling. وفى هذا الصدد فإن إجراء سحب عينة عشوائية خطية لا يتطلب سوى تحديد الإحداثيات على أحد المحورين، أو كليهما، عن طريق اختيار أرقامها من عمود واحد أو أكثر من جدول الأرقام العشوائية لتمثيل مفردات العينة المطلوبة. وعلى طول خط الإحداثى الذى حددناه عشوائياً يمكن قياس المسافات التى تشتمل - أو لا تشتمل - على خصائص معينة للظاهرة قيد البحث والتى تمثل قيمها مكوناً أساسياً للبيانات التى سيجرى عليها التحليل الكمى بعد ذلك.

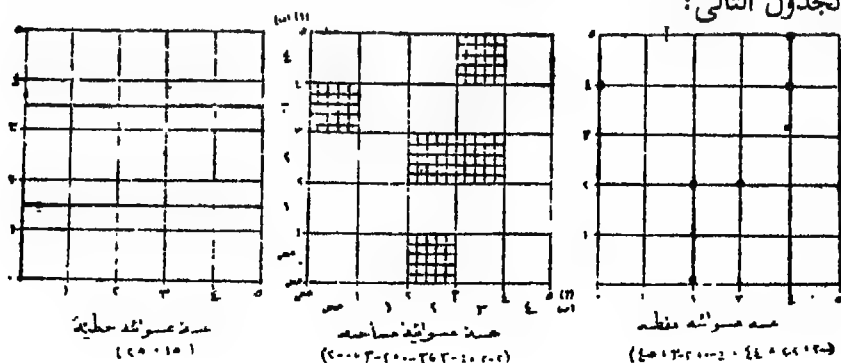
وجدير بالذكر أننا فى كل الحالات السابقة كنا نفترض أن المساحة التى تجرى عليها أحد طرق المعاينة تأخذ شكلاً مساحياً مربعاً أو مستطيلاً يمكن رسم شبكة متقنة ودقيقة من المربعات داخله، ولكن ليس هذا بالوضع الحقيقى فى الطبيعة لمعظم التوزيعات المكانية والتى قد تمثل مناطق أو تقسيمات إدارية تتخذ على الخريطة أشكالاً مساحية مختلفة تبعاً لتباين شكل حدودها. وللتغلب على هذه الصعوبة فإنه بإمكاننا استخدام مستطيل من ورق الشفاف، مرسوم عليه شبكة دقيقة من المربعات، تفوق مساحته المنطقة كلها أو يغطى معظمها على الأقل. وبذلك يمكن رفض أى إحدائى يختار من جدول الأرقام العشوائية ويقع بعيداً عن حدود المنطقة قيد البحث.

ح- طريقة السحب بواسطة الحاسب الآلى:

تستخدم هذه الطريقة فى حالة سحب عينات كبيرة من الحجم من مجتمعات تتميز بأحجام كبيرة جداً. وتحقق هذه الطريقة درجة عالية من العشوائية وعدم التحيز حيث أن الباحث لا يتدخل فى عملية الاختيار.

ويطبق هذا الأسلوب حالياً فى الأبحاث العلمية التى تجرى فى معظم دول غرب أوروبا والولايات المتحدة الأمريكية، وفى سبيله للإنتشار فى الدول الأخرى التى أخذت على عاتقها، حديثاً، تطوير أجهزتها الإحصائية بإدخال الحاسبات الآلية Computers ومن بينها جمهورية مصر العربية.

وبالإضافة إلى السحب الآلى لمفردات العينة، فإنه من الممكن الآن تغذية الحاسب الآلى ببرامج يمكن بواسطتها أن يضع جداول للأرقام العشوائية التى من أمثلتها الجدول التالى:



جدول رقم (١-٢) جدول الأرقام العشوائية بواسطة
الحاسب الآلى^(١)

٧	٣	٢	٤	٤	٧	٠	١	١	٤	٠	٧	٠	٤	٤	٠	١	٢	٤	٤	٥	٥	٩	٩	٠	٣	٣	
١	١	٣	٢	٤	٣	٩	٥	٢	٤	٠	٦	٢	١	٣	٢	٠	٩	١	٤	١	٣	٣	٧	٥	٧	٨	٠
٧	٦	٢	٤	٥	٢	٣	٢	١	٣	٤	٣	٧	٦	٨	٩	٨	٨	٥	٨	٤	٤	٦	٢	٣	٨	٣	٤
٣	٠	٩	٠	٧	٤	٣	١	٩	٢	٣	٣	٣	٧	١	٩	٥	٣	١	١	٤	٢	٨	٤	٤	٢	٨	٤
٢	٩	٢	٦	٢	٣	٤	٧	٢	١	٣	٨	٥	٦	١	٤	٢	٠	١	٢	٢	٨	٤	٦	٣	٠	٧	٢
٠	٤	١	٤	١	٩	٧	٧	٨	٦	٤	٣	٨	٠	٨	٩	٤	٤	٤	٧	٢	١	٢	١	٣	٣	٢	٢
٤	٣	٨	٦	١	٨	٤	٥	٣	٢	٤	٠	٣	٠	٤	٦	١	٨	٣	٧	٨	٣	٤	٣	١	٨	٥	٦
٢	٠	٣	٧	٠	٢	٧	١	٣	٩	٨	٨	٥	٨	٧	٩	٧	١	٧	٤	٠	٣	٠	٨	٤	٩	٦	١
٧	١	٤	٩	٦	٥	٥	٥	٩	٢	٩	٥	٦	٨	٢	٢	٧	٢	٤	٦	٣	٤	٦	٦	٧	٧	٨	٧
٩	١	٧	٢	٤	٩	٩	٤	٧	٣	١	٧	٤	٦	٣	٣	٨	٦	٧	٢	٠	٢	٩	١	٤	٥	٣	٩
٣	٨	٠	٤	٣	٠	٣	٥	٩	٧	٧	٤	٢	٠	٦	١	٥	٩	٩	٥	١	٢	١	٤	١	٩	٥	٤
٦	٩	٥	٦	٥	٤	٠	٤	١	٠	٩	٠	٩	٩	٥	١	٦	١	٥	٨	٦	٠	٠	٣	٨	٨	١	٨
٣	٧	٨	٥	٣	٦	٤	٠	٢	٢	٨	٤	٨	٩	٢	٩	٠	٣	٥	٤	٧	٩	٢	١	٥	٣	٧	٠
٠	٥	٦	٧	٧	٩	٥	٢	٩	٤	٢	٧	٤	٠	٢	٤	٣	٣	٢	٥	٣	٣	٧	٥	٠	١	٢	١
٣	٧	٩	٦	٥	٤	٣	٤	٩	٦	٠	٥	٨	٣	٠	٧	٥	٧	٦	٨	٩	٠	٧	٨	٣	٣	٩	٢
٩	١	٢	٤	٥	١	٠	٩	٤	٨	٥	٠	٤	٩	٧	٩	٩	٦	٠	٧	٠	٦	٤	٨	٩	٩	٤	٨

(١) وضع برنامج الحاسب الآلى للحصول على أرقام هذا الجدول Dr. M. McCullagh أستاذ
الكارتوجرافيا والتحليل الكمي فى الجغرافيا - بقسم الجغرافيا - جامعة نوتنجهام - إنجلترا.

7 7 7 2 2 2 . 2 . 7 1 . 0 9 1 7 8 2 2 . . 2 9 2 0 8 2
 1 8 9 7 7 8 8 8 7 9 0 9 9 9 7 7 . 0 0 1 7 0 7 7 7 7 2 7
 7 2 9 9 2 2 2 2 9 2 2 0 2 7 7 2 . 0 9 0 0 0 7 1 9 0 2 .
 7 0 7 7 9 7 7 2 7 7 7 7 7 2 2 9 1 1 . 7 8 7 7 8 7 8
 0 0 7 1 2 9 0 2 . 2 1 8 9 7 7 8 8 8 7 9 0 9 9 9 7 . 0 0
 7 8 7 7 1 7 8 7 8 7 7 2 9 9 2 2 2 2 9 2 2 0 2 7 2 . 0 9
 7 . 8 . 0 8 . 0 0 1 7 0 7 7 9 7 7 2 7 7 7 7 7 2 2 9 1 1
 2 0 7 2 2 9 1 1 2 8 7 8 0 8 9 2 7 7 7 . 2 . 7 1 8 2 0 2
 8 2 7 2 1 2 9 7 2 8 2 9 8 1 2 0 2 7 7 0 7 7 7 7 . 1 2 7
 9 2 7 . 8 1 9 2 0 9 2 0 0 7 0 9 8 . 7 0 7 1 2 7 8 . 7 .
 2 1 2 8 9 0 1 2 7 7 7 2 7 9 9 1 2 7 2 0 . 8 2 2 9 7 . 2
 8 2 . 2 7 2 7 7 2 7 7 2 . 2 . 7 1 2 2 2 2 9 2 9 2 2 7 2
 7 0 7 9 0 1 9 9 2 2 2 7 2 7 8 2 2 2 1 7 7 1 2 2 8 7 2 9
 9 9 . 1 7 0 2 9 2 2 . 2 1 2 2 7 2 2 2 9 2 0 7 1 0 . 9 7
 2 . 2 . 2 7 0 1 9 7 0 2 7 8 0 2 7 2 1 . 8 2 2 2 2 8 2 0
 8 2 7 8 1 1 9 7 7 9 7 2 1 2 2 2 2 2 9 . . 2 . 7 2 2 1 2
 0 . 7 7 0 0 7 2 8 . 8 2 1 7 0 9 7 1 0 2 2 0 2 9 9 7 1 8
 . 7 2 7 0 2 1 . 2 2 8 2 2 7 2 2 . 2 2 2 7 2 0 . . 7 8 1
 7 9 . 8 1 2 7 0 0 9 . 1 2 2 7 7 9 . 2 0 1 7 9 7 2 1 0 8

نخلص من كل مما سبق أنه عل بالرغم من أن الباحث يجب أن يكون حذراً وغير متحيزاً عند اختياره لمفردات العينة بإحدى طرق الاختيار السابقة، إلا أن هناك أنواع كثيرة من الأخطاء التي تصاحب أسلوب سحب العينات يكون مصدرها الرئيسي إما تحيز الباحث (خطأ التحيز Bias Error) أو تحيز المبحوث، أو عدم التزام الباحث بأسلوب العمل الإحصائي أى عدم اتباع القواعد السليمة فى جمع المعلومات أو سوء التقدير والإهمال فى العمل. وإلى جانب ذلك هناك أيضاً خطأ الصدفة أو الخطأ العشوائى Random Error الذى يعد من أهم أخطاء أسلوب المعاينة فى الدراسة. وكما سبق أن ذكرنا أن خطأ الصدفة من الأخطاء التى تخرج عن نطاق القصد أو التعمد حيث أن قيمة الخطأ تتفاوت من عينة لأخرى، وأن مصدره الأساسى هو الاعتماد على بيانات العينة فقط فى استخلاص النتائج الخاصة بالمجتمع الذى تمثله هذه العينة (راجع الفصل الأول).

مثال تطبيقي لكيفية سحب عينة:

لنفرض أنه لدراسة تباين الإنتاج الزراعى فى إحدى المحافظات عن طريق اختبار الفرق بين متوسطات إنتاج الأرض الزراعية فى قرى تلك المحافظة والتي يبلغ عددها ٩٥ قرية، ولما كان من الصعب دراسة القرى كلها لكثرة عددها أو لصعوبة الوصول إلى بعض منها، فقد تقرر أخذ عينة من ١٠ قرى. والمطلوب تحديد هذه القرى العشرة من بين القرى كلها. ولتحقيق ذلك تستخدم طريقة الجداول العشوائية لاستخراج أرقام القرى المطلوبة باتباع الخطوات التالية.

- ١- إعطاء كل قرية رقماً خاصاً بها من ١ إلى ٩٥.
- ٢- بالرجوع إلى جداول الأعداد العشوائية (جدول رقم ١-١) يمكن اتخاذ أى عمود أو صف واختيار عشرة أرقام منه.
- ٣- إذا أخذنا الصف الأول من الجدول فإن الأرقام العشوائية للقرى المختارة تكون هي:

٢٠، ١٧، ٤٢، ٢٨، ٢٣، ٥٩، ٦٦، ٣٨، ٦١، ٢ (لاحظ أننا لم نأخذ

الرقم السادس فى الصف وهو رقم ١٧ لأنه مكرر واستبدلناه بالرقم ٢ من نفس الصف).

٤- تكون الأرقام العشرة السابقة هى الأرقام العشوائية الممثلة لمجتمع القرى (٩٥).
أما إذا كان عدد القرى هو ٩٥٠ قرية وأريد اختيار عينة بحجم ما فإن أرقام مفردات العينة ستحتوى فى هذه الحالة على أرقام مكونة من ثلاثة حدود. فإذا كانت العينة مكونة من ٣٠ قرية مثلاً فإننا نأخذ العمودين الأول والثانى من جدول الأرقام العشوائية ونختار منها الأرقام المكونة من ثلاثة حدود، والتي تمثل مفردات العينة، فتكون هى:

٧٢٠، ٠٩٤، ٥٢٢، ٤٤٥، ١٤٤، ٣١٦، ٠٠٤، ٠٣٢، ٤٠٣، ٠٦١،
٣٨٩، ٢٠١، ٦٢٧، ٥٤٩، ٤٤٩، ٦٢٠، ٧٤٨، ٢٠٨، ٧٩٥، ٩٣٧، ٩٠٥،
٥٥٥، ٨٦٧، ٦٨٥، ٠٤٠، ٥٩٤، ٣١١، ٠٦٤، ٤٥٠، ٨٨٦.

وبلاحظ أننا رفضنا الأرقام ٩٧٤، ٩٩٣، ٩٦٢ نظراً لأنها أكبر من حجم المجتمع الأصلي (٩٥٠ قرية) واختارنا بدلاً منها الأرقام العشوائية الثلاثة الأخيرة: ٠٦٤، ٤٥٠، ٨٨٦.

(٣) تحديد نوع العينة:

يجمع كثير من الحصائيين والباحثين على أن تحديد نوع العينة المختارة، التى يجب أن تتوفر فيها صفة إعطاء نتائج ذات دقة معينة بأقل تكاليف ممكنة أو بأعلى دقة بتكاليف محدودة، يتوقف على طبيعة الدراسة، ونوعية وتركيب المجتمع الذى ستسحب منه العينة، والوسيلة أو الأداة المستخدمة فى جمع البيانات، ووجهة نظر الباحث نفسه.

ويمكن تصنيف العينات على أساس عامل العشوائية فى الاختيار إلى قسمين رئيسيين: القسم الأول يشمل العينات العشوائية التى يعتمد الباحث فى تصميمها على نظرية الاحتمالات فى إعطاء الفرص المتكافئة لمفردات المجتمع لأن تظهر فى العينة، أما القسم الثانى فيتضمن العينات العمدية (غير العشوائية) والتى يكون فيها

تحيز البحث واضحاً في اختيار مفردات العينة وذلك عن طريق إعطاء فرص غير متكافئة للمفردات نتيجة تعمد اختيار بعض المفردات دون غيرها من مفردات المجتمع الذي يريد معاينته. ولكل من القسمين أنواع متعددة ومتنوعة من العينات سندرسها بالتفصيل كما يلي:

أولاً: العينات العشوائية Random Samples

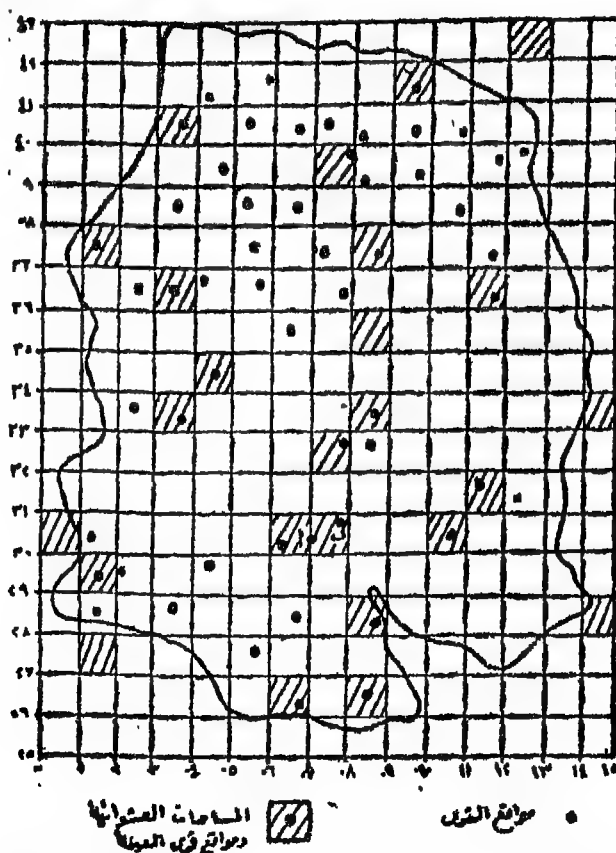
يستخدم أحياناً مصطلح العينات الاحتمالية للدلالة على العينات العشوائية وذلك لأنها تعتمد، كما سبق القول، على نظرية الاحتمالات في اختيار مفردات العينة من بين مفردات المجتمع عن طريق سحب تلك المفردات بالتتابع لكل منها احتمال معروف في الاختيار في السحبات المختلفة. وبعبارة أخرى تجرى المعاينة العشوائية (الاحتمالية) حسب خطة إحصائية لايسمح فيها للباحث أو حتى لمفردات العينة التدخل في اختيار العناصر الخاصة بالبيانات المراد جمعها أو أن يستعاض عن بعض المفردات التي يجب اختيارها وضمها للعينة بمفردات أخرى أسهل في حالة صعوبة الوصول إلى أو الحصول على بيانات المفردات الأولى. وسوف نعرض فيما يلي لأهم أنواع العينات العشوائية وهي: العينة العشوائية البسيطة، العينة العشوائية المنتظمة، العينة العشوائية الطبقية، والعينة العشوائية المتعددة المراحل.

أ- العينة العشوائية البسيطة Simple Random Sample

يلتزم هذا النوع من العينات الدراسات التي تهدف إلى تحديد خصائص المجتمع الذي تمثل مفرداته مجموعة متجانسة، أي ذات نوعية واحدة، مثل مجموعة من الطلاب عند مستوى عمرى متقارب وأوزان متساوية تقريباً، أو مثل إنتاج أحد مصانع المعلبات لوحداث إنتاجية (عبوات) ذات أوزان واحدة. ويتم اختيار مفردات العينة على أساس تساوى فرص الاختيار المستقل أمام كل مفردات المجتمع. مثال ذلك إذا أردنا اختيار عينة من ٢٠ قرية فى منطقة تحتوى على ٢٠٠ قرية لتكون فى قرى العينة صفات تمثل جميع القرى فى المنطقة، فإنه يجب اختيارها بالطريقة العشوائية التى عرفناها. كما يمكن بطريقة أخرى تحديد عدد

قرى العينة بتعيين مواقعها على الخريطة باستخدام الإحداثيات العشوائية (شكل رقم ١-٣). وفيما يلي أهم الخطوات التي تتبع لإجراء هذا الاختيار.

١- تبدأ بتحديد إطار المعاينة الذي يؤكد المسافات العشوائية بين القرى عن طريق استخدام مستطيل شبكة المربعات الذي يوضع فوق الخريطة، وتكون أماكن القرى المطلوبة ما يقع منها داخل المربعات التي تختار بطريقة الإحداثيات العشوائية. وفي هذه الحالة تقوم بترقيم الإحداثيات الرئيسية لشبكة المربعات، باختيار أرقام من جدول الأرقام العشوائية، ابتداء من الركن الجنوبي الغربي (نقطة الأصل) للخريطة. وبما أنه من المتوقع إختيار إحداثيات لمربعات خالية من مواقع القروى داخل المنطقة، أو مربعات يوجد بها قرى ولكنها تقع خارج



شكل رقم (١-٣): تحديد مواقع ٢٠ قرية باستخدام الاحداثيات العشوائية

حدود المنطقة قيد البحث، فإنه من المستحسن أن نأخذ عينة من ٢٥ مربعاً ثم نختار منها ٢٠ مربعاً تحتوى على مواقع القرى المطلوبة.

٢- بالرجوع إلى جدول الأرقام العشوائية (جدول رقم ١-١) نختار أى زوجين من الأرقام فى الجدول ليمثل كل زوج منها أحد الإحداثيين الشرقى أو الشمالى، وبذلك تكون الإحداثيات العشوائية لعدد ٢٥ قرية هى:

٠٧ - ٣٣	٠٦ - ٣٠	٠٨ - ٢٦	١٠ - ٣٠	١١ - ٣٦
٠٣ - ٣٦	١٤ - ٣٣	٠١ - ٢٩	١٤ - ٢٨	٠١ - ٢٧
٠١ - ٣٧	٠٣ - ٢٣	٠٣ - ٤٠	٠٨ - ٣٧	٠٧ - ٣٩
٠٦ - ٣٦	٠٨ - ٣٥	٠٤ - ٣٤	١١ - ٤٢	٠٩ - ٤١
٠٠ - ٣٠	٠٧ - ٣٠	٠٨ - ٣٣	٠٨ - ٢٨	١١ - ٣١

٣- يؤخذ الزوج الأول من الأرقام من كل عمود ليمثل الإحداثيات الشمالية لشبكة معاينة المربعات والتي يبدأ القياس على محورها من ٢٥ حتى ٤٣، بينما يمثل الزوج الآخر من الأرقام الإحداثيات الشرقية التى يبدأ القياس على محورها من صفر حتى ١٥.

٤- بطريقة الإحداثيات السابقة يمكن تحديد المربعات التى تقع بداخلها القرى المطلوبة. وتظل هذه المربعات حتى يمكن تمييزها عن بقية المربعات التى لم يقع عليها الاختيار. ويلاحظ من الشكل وجود خمسة مربعات مظلمة بدون أى موقع لقرية بداخلها. كما أنه فى حالة إذا ما وجد بداخل المربع المختار أكثر من موقع للقرى فيفضل اختيار موقع القرية القريب من الركن الجنوبي الغربى للمربع، بينما إذا كان هناك موقعاً لقرية فوق خط إحداثى يفصل بين مربعين، فيفضل ضم هذا الموقع إما إلى المربع شمال هذا الإحداثى أو إلى المربع فى شرقه. ويوضح الإحداثى (٣٠ - ٠٧) هذه الحالة والتى ضم فيها الموقع (أ) إلى المربع الذى يقع إلى الشرق من الإحداثى لأنه يقترب من الركن الجنوبي الغربى لهذا المربع، وأهمل الموقع (ب) على الرغم من وقوعه داخل المربع المختار بالإحداثى (٣٠ - ٠٧). ويجدر الإشارة إلى أن أساس مثل هذه القواعد يضعها الباحث بنفسه، كما يجب

أن نؤكد أنه ليس هناك أية خطورة من عنصر التدخل الشخصى للباحث فى ذلك طالما أن القواعد الموضوعية تتبع بطريقة ثابتة وبدون تغير (أو تبديل) على كل مفردات العينة.

ب- العينة العشوائية المنتظمة Systematic Random Sample

يستخدم هذا النوع من العينات عند دراسة المجتمعات التى تكون مفرداتها متخذة شكل انتظام متسق (أى تتصف بعدم التغير أو التباين الشديد) ومتدرج من حيث التنوع. وفى هذه العينة تأكيد على تسلسل المفردات وفقاً للتنوع فى الخصائص المميزة للمجتمع الأصلي. ويتبع فى اختيار مفردات العينة الأسلوب العشوائى، كما فى العينة العشوائية البسيطة، غير أن الاختيار يتم بطريقة منتظمة. فنقوم أولاً بترتيب مفردات المجتمع عشوائياً، أى بحيث تكون الفرص المتكافئة أمام جميع المفردات للظهور فى العينة هى أساس الاختيار، وعندئذ تنتهى العشوائية وبدأ اختيار مفردات العينة وفق نظام أو قاعدة معينة حتى نحصل على العينة المطلوبة.

وتمتاز هذه العينة بانتظام وثبات الفترات أو التباعد بين وحدات أو مفردات العينة. وفيها نبدأ بتحديد مقدار تمثيل كل مفردة من مفردات العينة لعدد معين من مفردات المجتمع، ثم نقوم بإختيار المفردة الأولى (أو أحد أعداد مقدار أو نسبة التمثيل) عشوائياً ونضيف إليها مقدار التمثيل بطريقة منتظمة حتى نحصل على بقية مفردات العينة بشكل منتظم وتباعد متساوى فيما بينها. ولتحديد مقدار التمثيل نقسم عدد مفردات المجتمع الأصلي على حجم العينة المطلوب. فمثلاً إذا كان حجم المجتمع الأصلي ٦٠٠٠ مفردة، وأردنا إختيار عينة من ٢٠٠ مفردة، فهذا يعنى أننا نريد إختيار مفردة واحدة لكل ٣٠ مفردة (أى $\frac{6000}{200} = 30$).

وإحدى الطرق المتبعة هى: أن نختار عدداً عشوائياً بين ١، ٣٠. ولو فرضنا أن هذا الرقم هو ٢٢، فإنه يمكن تحديد مفردات العينة مباشرة بإضافة مقدار التمثيل (٣٠) بطريقة منتظمة إلى الرقم (٢٢) وذلك على النحو التالى:

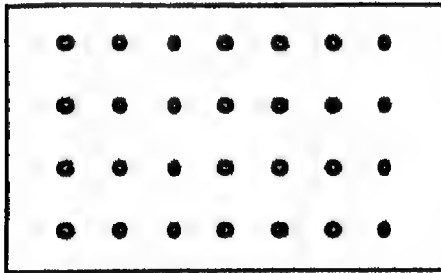
٢٢، ٥٢، ٨٢، ١١٢، ١٤٢، ١٧٢، ٢٠٢، ٢٣٢، ٢٦٢، ٢٩٢ ... وهكذا حتى نصل إلى المفردة الأخيرة ويكون رقمها ٥٩٩٢. وهناك طريقة لإيجاد ترتيب أى مفردة من مفردات العينة وذلك باستخدام المعادلة الآتية:

ترتيب المفردة = الرقم العشوائى المختار + (حجم العينة - ١) × مقدار التمثيل
أى أن:

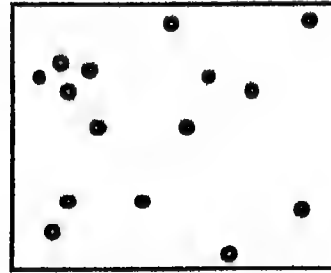
فإذا أردنا مثلاً معرفة ترتيب المفردة رقم ١٠٠ من ٢٠٠ مفردة فى العينة السابقة فإن.

$$٢٩٩٢ = (٣٠ \times (١ - ٢٠٠) + ٢٢) = ١٠٠$$

ويمكن تطبيق العينة المنتظمة فى مجال الدراسة الجغرافية، إذ يختار موقع المفردة الأولى عشوائياً ويوقع على الخريطة بالنسبة لكل من إحداثياتها الشرقى والشمالى من نقطة الأصل للخريطة، ثم تختار مواقع بقية المفردة الأخيرة بشكل منتظم يمثل تباعد (مسافة) بالنسبة للمحورين العشوائيين الأفقى والرأسى كما فى الشكل رقم (١-٤)



٢- عينة عشوائية منتظمة



١- عينة عشوائية بسيطة

شكل رقم (١ - ٤): التوزيع المكاني لمفردات العينة العشوائية

وتتميز العينة المنتظمة بأنها أسهل وأسرع فى التطبيق من العينة العشوائية البسيطة، إذا أنها لا تحتاج إلى اختيار كل المفردات بالطريقة العشوائية والذي قد ينتج عنه تكرار سحب بعض المفردات. كما أنها تمثل توزيعاً متسقاً (منتظماً) Uniform للمجتمع الذى تسحب منه العينة بعكس العينة العشوائية البسيطة التى

ينتج عنها فى معظم الأحيان توزيعات مكانية تتخذ أشكالاً عنقودية Clusters or Bunching تتباعد فيها المفردات بعضها عن بعض تباعداً لا يمكن تقليله إلا بزيادة حجم العينة. فمثلاً إذا أراد أحد الباحثين تقدير نسبة الأراضى الزراعية من خريطة منطقة ما (شكل رقم ١ - ٤)، وكانت أكبر المساحات للأراضى الزراعية تتركز بالقرب من المكان (أ) فإن قيمة النسبة فى هذه الحالة ستكون أعلى من القيمة الحقيقية لها.

ويعاب على العينات المنتظمة إذا أجريت على التوزيعات المكانية أنها تؤكد صفة الانتظام والاتساق فى الظاهرة أو الظاهرات الجغرافية موضع المعاينة والتي تكون غالباً، إن لم يكن دائماً، فى حالة تغير وتطور مستمرين. كما قد تتصف المعاينة المنتظمة بأنها لا تعطى عينة غير متحيزة أو ممثلة للمجتمع الذى سحبت منه بسبب خطأ التحيز وعدم اتباع أسلوب تكافؤ الفرص فى اختيار مفردات العينة إلا إذا كانت مفردات المجتمع الأصلية موزعة توزيعاً عشوائياً - وكثير من المجتمعات الاحصائية للظواهرات الجغرافية لا تتصف بصفة العشوائية فى توزيعاتها. وعلى العموم فإن خطأ التحيز إن وجد فى بيانات العينة المنتظمة فإن قيمته تكون صغيرة جداً بحيث يمكن إهمالها عند تطبيق أساليب التحليل الإحصائى الكمى على هذه البيانات.

ج- العينة العشوائية الطباقية Stratified Random Sample

يستخدم هذا النوع من العينات العشوائية فى دراسة المجتمعات التى تتميز بتباين نوعيات مفرداتها والتي يمكن تقسيمها إلى مجموعات (أو طبقات Strata) لكل مجموعة أو طبقة منها خصائص ومميزات معينة، ولكنها تتباين فيما بينها تبايناً عظيماً فى هذه الخصائص والمميزات. كما يتسم هذا النوع من العينات بأنه أكثر دقة فى الاختيار العشوائى من العينات العشوائية البسيطة حيث أن المفردات الكلية للعينة الطباقية تكون أكثر تمثيلاً لجميع مجموعات أو طبقات المجتمع الأصلية، مما يؤدي إلى تقليل فى الأخطاء عند تعميم نتائج العينة على كل المجتمع. وهكذا تقوم العينة الطباقية على أساس تقسيم المجتمع الأصلية إلى طبقات ثم نأخذ عينة

عشوائية من كل طبقة بشكل يتناسب مع مفردات أو حجم تلك الطبقة. ولهذا فإن معاينة أية طبقة تتطلب عدة إجراءات تختلف عن تلك التى تتطلبها الطبقة (أو الطبقات) الأخرى. فمثلاً فى التعدادات بالعينة التى تجرى لحصر أعداد السكان فى منازل أحد الشوارع بمدينة ما تختلف عن مثيلتها لحصر عدد العاملين فى مؤسسة أو شركة ما.

وبصفة عامة يمكن تلخيص طريقة اختيار العينة الطباقية فى الخطوات الآتية:

١- تقسيم المجتمع إلى مجموعات مميزة أو فئات فرعية (مجتمعات صغيرة) متجانسة تعرف بالطبقات Strata.

٢- تحديد نسبة مفردات كل مجموعة أو طبقة Stratum بالنسبة للعدد الكلى لمفردات المجتمع الأصلى (أى $\frac{\text{حجم الطبقة}}{\text{حجم المجتمع الأصلى}}$).

٣- تحديد عدد مفردات العينة المطلوبة من كل طبقة أو ما يعرف بالعينة الفرعية Subsample التى تتحدد عن طريق نسبة حجم كل طبقة فى المجتمع الأصلى والحجم الكلى للعينة.

٤- استخدام الأسلوب العشوائى لاختيار المفردات من كل طبقة.

مثال: أراد باحث سحب عينة حجمها ٥٠٠ عاملاً من مجتمع عمالى لأحد الصناعات يبلغ حجمه ٥٠٠٠ عاملاً وذلك حسب الحالة التعليمية، فقام بتقسيم عمال المجتمع إلى فئة من العمال الأميين وعددهم ١٠٠٠ عاملاً، وفئة من العمال الحاصلين على شهادات أقل من المتوسطة ١٥٠٠ عاملاً، وفئة من العمال الحاصلين على شهادات متوسطة وفنية وعددهم ٢٥٠٠ عاملاً. فكم من العمال من كل فئة يمكن اختيارهم حتى يحصل على الحجم الكلى للعينة، ويتم ذلك بتحديد حجم العينة الفرعية عدد المفردات المطلوبة المثلة لمجتمع الطبقة وهو يساوى:

$$\text{حجم العينة} \times \frac{\text{حجم الطبقة}}{\text{حجم المجتمع الأصلى}}$$

بالتالى فإن:

$$\text{عدد المفردات من الفئة (عمال أميين)} = \frac{1000}{5000} \times 500 = 100 \text{ رطلاً}$$

$$\text{عدد المفردات من الفئة (عمال الشهادات أقل من المتوسطة)} = \frac{1500}{5000} \times 500 = 150 \text{ عاملاً}$$

عاملاً

$$\text{عدد المفردات من الفئة (عمال الشهادات المتوسطة والفنية)} = \frac{2500}{5000} \times 500 = 250 \text{ عاملاً}$$

عاملاً

ويكون عدد العمال المطلوب اختيارهم من كل فئة من الفئات الثلاث هو على الترتيب ١٠٠، ١٥٠، ٢٥٠ ومجموعهم لابد وأن يساوى الحجم الكلى للعينة المطلوبة ٥٠٠ عاملاً من المجتمع الأصلي لعمال تلك الصناعة.

وتجدر الإشارة هنا إلى أن الطبقة يمكن أن تكون طبقة طولية أو عرضية (فى حالة مجتمع المساحات)، أو طبقية أفقية أو رأسية (فى حالة نوع المناطق السكنية أو مستويات الدخل أو فئات العمر). وفى مجال الدراسات الجغرافية تستخدم المعاينة الطبقة بصفة خاصة عند دراسته أو مقارنة بعض المظاهر الجغرافية لمنطقتين أو أكثر، أو عند مقارنة بيانات متغيرين (ظاهرتين) وذلك عندما يكون تباينهما غير موزع بنسب متساوية بين أرجاء المنطقة قيد البحث. وفى الحالة الأولى مثلاً إذا أراد الباحث دراسة عينة من الأحواض الزراعية بأحد المراكز الهامشية للدلتا وذلك ضمن دراسة الخصائص الطبيعية والاقتصادية للتربة، فإن ذلك يستلزم منه تقسيم أحواض المركز حسب نوع التربة إلى أنواع أو طبقات مميزة من التربة (تربة رملية، تربة طينية، تربة طينية)، واختيار عينات فرعية من الأحواض الزراعية متساوية العدد من كل من نوع من أنواع التربة، بينما فى الحالة الثانية إذا أريد فحص العلاقة بين ظاهرة (أو متغير) نمو النبات الطبيعى ومستوى الارتفاع عن سطح البحر لإحدى المظاهر التضاريسية الوعرة، فإن ذلك يستلزم من الباحث أن يختار عينة مناسبة (٢٠ مكاناً مثلاً) ومثلة لكل مستوى من مستويات (نطاقات) الارتفاع

(وهي عبارة عن المساحات بين خطوط الكنتور) على حدة. على أن تجمع البيانات الخاصة بكل متغير من المتغيرين: نمو النبات ومستوى الارتفاع من أماكن العينة. ويتم ذلك عن طريق تقسيم منطقة الدراسة إلى نطاقات (مستويات) ارتفاع تنحصر بين خطوط الكنتور: صفر - ١٠٠ متر، ١٠٠ - ٢٠٠ متر، ٢٠٠ - ٣٠٠ متر وهكذا، ثم يختار بعد ذلك ٢٠ مكانا داخل نطاق أو مستوى الارتفاع الأول اختيارا عشوائيا مع رهمال الأماكن التي تزيد عن حجم العينة الفرعية داخل هذا النطاق. ويستمر إجراء نفس العمل لكل نطاق حتى تكتمل بقية النطاقات ويصبح بداخل كل منها عدد المفردات ٢٠٠ مكانا المطلوب لمعاينة المنطقة. وتعرف العينة في هذه الحالة بالعينة الطبقيّة المنتظمة Stratified Systematic حيث أنه قد تم فيها اختيار عدد مفردات متساوي في كل الطبقات.

أمثلة تطبيقية:

مثال (١) لدراسة أحد مظاهر النشاط الصناعي مثل صناعة الأحذية في منطقة تتميز بتباين توزيع منشآت أو ورش هذه الصناعة في مختلف المراكز العمرانية في هذه المنطقة ، فأنا نقوم بمعاينة توزيع هذه الصناعة بطريقة المعاينة الطبقيّة عن طريق تقسيم المراكز العمرانية في المنطقة حسب حجم السكان بها، إلى أربع مجموعات: قرى صغيرة، قرى كبيرة، مدن صغيرة، مدن كبيرة، ويختار عينة عشوائية من المراكز العمرانية لكل مجموعة بنسبة معاينة $\frac{1}{10}$ أي $\frac{\text{حجم العينة}}{\text{حجم المجتمع}} = \frac{n}{N}$ والتي تتناسب وحجم كل مجموعة من المجموعات الأربع، كما يتناسب المجموع الكلي للعينات مع الحجم الكلي لمجتمع المراكز العمرانية. وتوضح ذلك البيانات التالية:

جدول رقم (١ - ٣)

طريقة حساب عينة طبقية من المراكز العمرانية

أ- عدد الوحدات (المراكز العمرانية) في العينة	ب- العدد الكلى للوحدات فى كل مجموعة	ج- عدد ورش صناعة الأحذية فى كل وحدات العينة	د- متوسط ورش العينة لكل وحدة	هـ- تقدير المجموع الكلى لعدد الورش فى كل مجموعة (ب × د)	المجموعة
١٢	١٢٠	٢٤	$\left[\frac{2}{1} \right]$	٢٤٠	قرى صغيرة
١٤	١٤٠	٣٥	٢,٥	٣٥٠	قرى كبيرة
٥	٥٠	٦٠	١٢	٦٠٠	مدن صغيرة
٢	٢٠	٨٠	٤٠	٨٠٠	مدن كبيرة
٣٣	٣٣٠	١٩٩	٦,٠٣	١٩٩٠	المجموع

من البيانات السابقة الخاصة بعينة طبقية مكونة من ٣٣ مركزاً عمرانياً من العدد الكلى المقدّر بنحو ٣٣٠ مركزاً عمرانياً يمكن تقدير متوسط عدد ورش صناعة الأحذية فى كل مجموعة من المجموعات الأربع، وتقدير متوسط عدد الورش فى كل مركز عمرانى على حدة، بالإضافة إلى تقدير العدد الكلى لهذه الورش التى توجد فى المراكز العمرانية لكل مجموعة على اختلاف أحجامها وفى كل المنطقة موضع الدراسة. وبالتالي فإنه عن طريق استخدام مثل هذه العينات التى تتكون من مجموعة صغيرة نسبياً من المراكز العمرانية يمكن إعطاء صورة تفصيلية عن توزيع الظاهرة قيد البحث، وتعميم نتائج التحليل الإحصائى الكلى لهذا التوزيع على جميع المراكز العمرانية فى كل أرجاء المنطقة.

مثال (٢): لدراسة تأثير طول فترة الإقامة فى المنطقة التجارية لمدينة ما على مفهوم المركز التجارى للمدينة لدى السكان وتصورهم ذهنى لهذا الجزء من

المدينة، فإن خطة الباحث في ذلك تنحصر في إجراء سحب عينة تكون ممثلة لجميع السكان بقدر المستطاع لقياس هذا التأثير. ويحتاج الباحث قبل إجراء عملية المعاينة أن تتوفر لديه بعض المعطيات الإحصائية عن خصائص سكان تلك المنطقة التجاريتي يمكن الحصول عليها من التعدادات السكانية أو غيرها من المصادر. فمثلاً قد تتجمع لدى الباحث بيانات تفيد أن سكان المنطقة التجارية في المدينة موضع الدراسة والذين تنحصر أعمارهم بين ٣٠ و ٥٠ سنة ينقسمون من حيث مدة الإقامة في المنطقة إلى ثلاث مجموعات: المجموعة الأولى تشمل السكان الذين أقاموا طوال حياتهم في المنطقة ونسبتهم ٦٠٪ من جملة السكان، والمجموعة الثانية تتكون من سكان أقاموا في المنطقة لمدة تتراوح من ٥ إلى ١٠ سنوات ونسبتهم ٢٠٪ من جملة السكان، والمجموعة الثالثة تتضمن أقاموا في المنطقة لمدة ٥ سنوات أو أقل ونسبتهم ٢٠٪ من المجموع الكلي للسكان. وفي هذه الحالة فإنه لجمع المعلومات اللازمة لمثل هذه الدراسة، يقوم الباحث بإجراء مقابلة لعينة من السكان من كل مجموعة، كأن يختار مثلاً ١٢٠ ساكناً في نفس فئة العمر (٣٠ - ٥٠ سنة) من الذين أقاموا في المنطقة طوال حياتهم، ٤٠ ساكناً في نفس فئة العمر أيضاً من بين الذين عاشوا في المنطقة لمدة تتراوح بين ٥ سنوات و ١٠ سنوات، ٤٠ ساكناً من الذين أقاموا لمدة ٥ سنوات أو أقل في المنطقة. وبإجراء ذلك فإن الباحث يكون قد قسم مفردات العينة على أساس مدة الإقامة إلى ثلاث طبقات يتناسب حجم مفردات كل منها مع الحجم الأصلي لنفس الطبقة من المجتمع. أما إذا أراد الباحث عمل مقارنة بين الطبقات الثلاث فإنه يقوم بأخذ ثلاث عينات فرعية Sub-Samples متساوية الحجم حتى تكون المقارنة سليمة. ويعاب على الطريقة الأخيرة عند تطبيق المعاينة الطباقية أن المجموع الكلي للمفردات المختارة قد لا يكون ممثلاً Representative لجميع مفردات الظاهرة قيد البحث.

مثال (٣): في دراسة لاستخدام الأرض وعلاقته بمظاهر السطح في منطقة ما، أخذت عينة عشوائية من ١٠٠ موقع من تلك المنطقة، ثم صنف هذه المواقع بحسب فئات الاستخدام المختلفة (أراضي زراعية، حشائش، غابات، أراضي جرداء)

وعلى أساس مستويات (طبقات) الارتفاع، المتباينة، فكانت بيانات هذه العينة كالآتي:

مستويات الارتفاع (متر)	أراضي زراعية	حشائش	غابات	أراضي بور	حجم العينة
أقل من ١٥٠	٣	٤	١	-	٨
١٥٠ - ٣٠٠	٥	٢١	٥	٣٠	٤١
أكثر من ٣٠٠	-	٦	-	٤٥	٥١
المجموع	٨	٣١	٦	٧٥	١٠٠

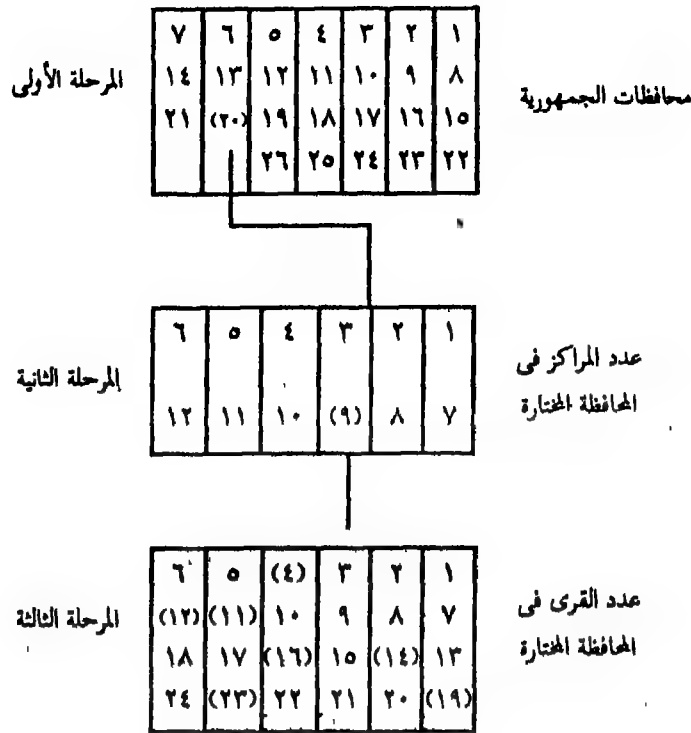
وتتطلب مثل هذه الدراسة حساب الخطأ المعياري لفئات استخدام الأرض في كل طبقة أو مستوى من مستويات الارتفاع وللعينة ككل. ويحسب الخطأ المعياري لمثل هذه البيانات (بيانات نسبية) على أساس أنه يساوي الجزء التريبيعي لحاصل ضرب نسبة وجود الظاهرة (أى فئة من فئات استخدام الأرض) فى نسبة عدم وجودها فى حجم العينة. ولتأخذ فئة الاستخدام الثانية (أراضي الحشائش) لتوضيح ذلك كما يلى:

مستويات الارتفاع (متر)	تكرار أراضي الحشائش	حجم العينة ن	١- نسبة وجود أراضي الحشائش فى العينة	ب- نسبة عدم وجود أراضي الحشائش فى العينة	الخطأ المعياري = $\sqrt{ن \times ب \times \frac{ب}{ن}}$
أقل من ١٥٠	٣	٤	١	-	٨
١٥٠ - ٣٠٠	٥	٢١	٥	٣٠	٤١
أكثر من ٣٠٠	-	٦	-	٤٥	٥١
المجموع	٨	٣١	٦	٧٥	١٠٠

ويمكن بعد ذلك تحويل الخطأ المعياري في كل مستوى (طبقة) من مستويات الارتفاع وللعينة كلها إلى نسبة مئوية ($\frac{\text{الخطأ المعياري}}{\text{حجم العينة}} \times 100$). فيمكن مثلاً التعبير عن التكرار المطلق لمواقع العينة الفرعية لمستوى الارتفاع ١٥٠ - ٣٠٠ متراً (٢١ موقعاً من جملة المواقع لهذا المستوى وهي ٤١ موقعاً وبخطأ معياري ٣٧٢, ٩) بتكرارات أخرى نسبية، كأن نقول أن التكرار النسبي الحقيقي لمواقع هذه الفئة (أراضي الحشائش) من فئات استخدام الأرض في المنطقة في مستوى الارتفاع ١٥٠ - ٣٠٠ متراً بدرجة ثقة أو عند المستوى الاحتمالي ٩٥٪ يمكن أن ينحصر بين ٥, ٤٨٪ و ٩٦, ٩٢٪.

د- العينة العشوائية المتعددة المراحل Multi-Stage Random Sample

يلتزم هذا النوع من العينات العشوائية دراسة المجتمعات الكبيرة، مثل الدراسات السكانية أو الدراسات في مجال الجغرافية الاقتصادية، وهي مجتمعات يمكن تقسيمها إلى عدد من الأقسام المتشابهة التي يحتوى كل قسم منها على عدد من المفردات التي تتصف بعدم التجانس في خصائصها، وإذا أطلق على هذا النوع من العينات بأنه «متعدد المراحل». فمثلاً إذا أردنا دراسة الحالة الاجتماعية في الريف على مستوى محافظات الجمهورية، فإننا نقوم أولاً باختيار عشوائي لعدد من محافظات الجمهورية، وبعد ذلك في مرحلة تالية نقوم باختيار عشوائي لعدد من مراكز المحافظات المختارة سابقاً، ثم تأتى بعد ذلك المرحلة الثالثة وفيها نقوم باختيار عشوائي لعدد من قرى المراكز المختارة في المرحلة الثانية، وتكون هذه القرى بما يختار منها عشوائياً من أسر عبارة عن المفردات التي تجرى عليها لتحديد بعض المؤشرات والمقاييس الاحصائية. ويهدف التدرج السابق في أخذ العينات في مراحل إلى التبسيط والحفاظ على طبيعة المفردات غير المتجانسة داخل العينة في كل مرحلة من المراحل. ويوضح ذلك الشكل التالي.



ويعاب على المعاينة المتعددة المراحل أن كثرة عدد المراحل التي قد تتضمنها تضعف العلاقة بين معالم المجتمع الأصلي وخصائص العينة مما يؤثر بالتالي على تقدير معالم المجتمع من بيانات العينة المتحصل عليها في آخر مرحلة، كما أن هذا النوع من العينات يتطلب من الباحث جهداً كبيراً في تحديد حدود أو إطار كل مرحلة وتحديد حجم العينة الفرعية المطلوبة من كل منها وذلك في ضوء الاعتبارات الخاصة بالاختيار في المعاينة العشوائية.

ثانياً: العينات غير العشوائية (العمدية) Non-Random Samples

كثيراً ما يتعرض أسلوب المعاينة العشوائية لبعض العقبات التي تحول دون التمسك به أو الاعتماد عليه في دراسة المجتمعات، وذلك عندما يتطلب سحب العينة العشوائية امكانيات مادية وفنية، أو عندما يجد الباحث صعوبة في الوصول إلى وحدة من وحدات المجتمع المختارة، أو في حالة عدم معرفة كل مفردات المجتمع الذي ستسحب منه العينة العشوائية. وفي مثل هذه الحالات يضطر الباحث إلى

اتباع أسلوب التعمد والتحيز الشخصى فى اختيار مفردات العينة، أو ما يعرف بأسلوب العينة العمدية (غير الاحتمالية). وبذلك يقوم اختيار هذا النوع من العينة على أساس شخصى ولا تراعى فيه الفرص المتكافئة للمفردات لأن تكون ضمن العينة، أى لا تراعى فيه صفة العشوائية.

وكثيراً ما تستخدم المعاينة العمدية، بصفة عامة، فى الأبحاث الاستطلاعية. كما فى حالة تقدير معالم مجتمع كبير، أو عند محاولة معرفة فكرة تقريبية سريعة عن خصائص ظاهرة ما بحيث لا تستخدم نتائجها للتعميم على المجتمع. كما تستخدم المعاينة العمدية فى الاختبارات القبلىة (السابقة) Prior Tests مثل اختبار صحيفة الأسئلة لمعرفة مدى تجاوب المبحوثين حتى يمكن إجراء التعديلات اللازمة فى الأسئلة قبل بدء المعاينة الرئيسية، أو فى حالة القيام ببعض القياسات لظاهرة ما لمعايرة الأجهزة المستخدمة فى القياس والتأكد من سلامتها. وجدير بالذكر أنه لا توجد هناك أية طريقة احصائية لمعرفة وقياس دقة نتائج المعاينة العمدية غير الاحتمالية، ولذا لا تعد هذه الطريقة من طرق المعاينة الجيدة، إلا أنه فى بعض الأحيان قد لا يجد الباحث سبيلاً مكنياً عملياً للمعاينة سوى استخدام هذه الطريقة. وسنعرض فيما يلى لأهم أنواع العينات غير العشوائية (العمدية أو غير الاحتمالية) وهى العينة الغرضية، العينة بالحصة، والعينة العنقودية.

أ- العينة الغرضية:

تلائم طريقة العينة الغرضية الدراسات التى تخص الظواهر التى تشتد فيها درجة تباين متعاطراتها، مما يجعل الباحث مضطراً إلى تحديد واختيار المتغيرات الخاصة بالبيانات المراد جمعها والتى يرى من وجهة نظره أنها تصلح للدراسة. فمثلاً الباحث الذى يدرس مستوى المعيشة فى الريف المصرى لا يمكنه الاعتماد على الاختيار العشوائى لعينة من القرى، بل يعتمد على تحديد عدد من القرى تمثل فى نظره مجتمع القرى المصرية وتكون محلاً للدراسة.

ب- العينة بالحصة Quota Sample:

يضطر الباحث إلى استخدام مثل هذا النوع من العينات العمدية (غير الاحتمالية) عندما يتطلب منه القيام بإجراء عدد معين من المقابلات لأشخاص لهم صفات محددة في مكان معلوم، أو بإجراء عدد معين من الزيارات الميدانية لجمع بيانات عن ظاهرة معينة داخل منطقة محدودة. وفي طريقة العينة بالحصة لا تختار مفردات (وحدات) العينة عشوائياً ولكن تستخدم أية معلومات تساعد على الحصول على الحصة المطلوبة بسرعة وتكاليف قليلة. ولذلك فإن هذه الطريقة تستخدم بكثرة في معاينة واستطلاع الرأي العام كما هو متبع في معهد جالوب بالولايات المتحدة الأمريكية عند التنبؤ بنتيجة الانتخابات العامة، إذ يطلب من الباحثين في هذه الحالة التعرف على رأى مجموعة من الناخبين على أن تكون من بينهم نسبة معينة من فئات مختلفة مثل فئات أصحاب المهن الحرة، وفئة العمال وفئة الموظفين... إلخ، ويترك للباحثين حرية التصرف في اختيار الأعداد المطلوبة منهم بأية طريقة يجدونها سهلة ومناسبة.

وقبل إجراء العينة بالحصة يجب التأكد من مجموعة الخصائص (ثلاث أو أربع خصائص مثلاً) التي تميز المجتمع الأصلي بحيث ترتبط هذه الخصائص ارتباطاً وثيقاً بالمتغير قيد البحث، وتصمم عينة تكون ممثلة لهذه الخصائص مجتمعة. ويتضمن تصميم العينة بالحصة ثلاث مراحل هي:

- ١- مرحلة تصنيف المجتمع الأصلي على أساس الخصائص موضع الدراسة.
 - ٢- مرحلة تحديد نسبة المجتمع في كل طبقة أو فئة.
 - ٣- مرحلة تحديد الحصص التي يراد دراستها واختيارها في ضوء العدد المطلوب.
- وجدير بالذكر أنه يمكن اعتبار العينة بالحصة نوع من العينات الطبقية التي يكون فيها الاختيار داخل الطبقة إختيار غير عشوائي، مما قد يؤدي إلى الوقوع في خطأ التحيز من قبل الباحث من جراد التصنيف الشخصي للعناصر والفئات من ناحية وعدم عشوائية الاختيار من ناحية أخرى.

ح- العينة العنقودية Cluster Sample :

تشبه طريقة المعاينة العنقودية للعينة متعددة المراحل فى كثير من مراحل إجرائها. وتقوم هذه الطريقة على أساس إختيار مفردات العينة فى حزم أو عناقيد Clusters بأقل جهد وتكلفة مما يحدث بالنسبة للعينة العشوائية. فمثلاً إذا كنا بصدد دراسة مستوى المعيشة فى منطقة متخلفة بأحد الأقسام الإدارية فى محافظة الإسكندرية، وبافتراض عدم وجود سجل يضم سكان هذه المنطقة، ولكنهم يوجدون فى سجلات مصلحة الكهرباء للمحافظة كلها بما فيها المنطقة قيد البحث، فإنه يمكن إختيار العينة على عدة مراحل أو تدريجياً بشرط أن تكون متكاملة.

ولإجراء سحب عينة عنقودية، نفترض أن دراسة جغرافية اجتماعية (مثل دراسة خصائص النشاط السكانى والعوامل المؤثرة فيه) ستجرى على سكان مدينة ما يبلغ تعدادها ٣٠٠٠٠ نسمة والمسجلين فى قوائم يمكن الحصول عليها بسهولة. والمطلوب أن تكون العينة التى تجرى عليها لدراسة مكونة من ٣٠٠٠ شخص فقط. وفى هذه الحالة نختتم طريقة العينة العنقودية أن تكون العينة متركزة (أو متجمعة) فى أجزاء قليلة من المدينة. فإذا افترضنا أن المدينة تنقسم إلى ٥٠ شياخة فى كل منها ٦٠٠ شخصاً فإنه يمكن إختيار عينة من خمس شياخات فقط (أى شياخة واحدة لكل ١٠ شياخات) وتجرى لدراسة على هذه الشياخات الخمس بما تحتوى من سكان.

من العرض السابق لكل من أسلوبى الحصر الشامل والمعاينة يتضح أن أسلوب الحصر الشامل يتعرض لخطأ التحيز الذى ينتج عن أى تقصير أو إهمال فى خطوات البحث، بينما يتعرض أسلوب المعاينة لخطأ الصدفة أو الخطأ العشوائى، وهو خطأ يصاحب عملية الاختيار نفسها. كما يتعرض أسلوب المعاينة لخطأ التحيز إذا حدث أى خطأ أو تقصير فى خطوات البحث بالعينة. ولكن تفضل الدراسة بالعينة كثيراً على الدراسة بالحصر الشامل لأن العينة كجزء من المجتمع تعطى فرصة كبيرة للتحكم والسيطرة على جميع العوامل المحيطة أو الخارجية. كما يمكن استخدام الخطأ العشوائى للعينة فى تقدير مستوى الثقة التى تعمم بها النتائج على المجتمع من بيانات هذه العينة.

طرق (أدوات) جمع البيانات:

بعد أن يتم تحديد الأسلوب الذى على أساسه سيتم جمع البيانات سواء كان أسلوب الحصر الشامل أو أسلوب العينات، فإنه يجب اتباع إحدى الطرق أو الوسائل (الأدوات) الاحصائية التى تستخدم فى عملية الحصول على البيانات لإتمام العمل الاحصائى لأى من الأسلوبين. وأهم طرق أو أدوات جمع البيانات من مصادرها: طرق المراسلة والاتصال والعمل الحقلى (الميدانى)، وتنقسم كل أداة منها إلى مجموعة من الوسائل يتم بواسطتها إتصال الباحث بمفردات المجتمع أو العينة. ولكل أداة مزايا متعددة فى استخدامها بما يتفق مع طبيعة البحث. فالمراسلة والاتصال أكثر الأدوات كفاءة فى التعداد والبحوث والدراسات السكانية، والعمل الحقلى أو المسح الميدانى بما يتضمن من وسيلة الملاحظة أو المشاهدة أكثرها كفاءة فى دراسة سلوك الظواهر الجغرافية الاجتماعية أو فى الكشف عن تفاصيل الظواهر وعن الفصائل الخفية التى توجد بين عناصرها أو بينها وبين بعض الظواهر الأخرى.

وجدير بالذكر أن اختيار أى أداة من أدوات جمع البيانات يتوقف على طبيعة المعلومات التى يراد جمعها والوقت المسموح به والامكانيات المادية المتاحة للباحث.

أولاً: المراسلة والاتصال:

يعتمد الباحث على هذه الأداة فى جمع البيانات إذا تعذر الوصول أو الاتصال المباشر بمفردات المجتمع أو العينة. وتتم عملية جمع البيانات بهذه الأداة عن طريق إرسال استمارة البيانات الاحصائية بالبريد، أو بواسطة الاتصال التليفونى بالمبحوثين، أو حتى عن طريق نشر الأسئلة على صفحات المجلات أو الجرائد المتخصصة.

أ- المراسلة بالبريد: يقوم الباحث فى هذه الطريقة بإرسال رسالة للشخص الذى وقع عليه الاختيار لاستبيانته يوضح له فيها أهمية البحث وأهدافه وسرية البيانات وعدم استخدامها إلا لغرض البحث فقط، ومع الرسالة يرفق الباحث الاستمارة الاحصائية المطلوب الإجابة على أسئلتها كما يرسل مع الرسالة مظروف

خاص بعنوان الباحث وخالص الرسوم البريدية، ويطلب من المبحوث إعادة الاستمارة الإحصائية مرة أخرى إلى الباحث.

ومن يرى هذه الطريقة أنها سهلة التنفيذ وقليلة التكاليف، إلى جانب أنها تعطى فرصة كافية للمبحوث فى التفكير والإجابة على الأسئلة دون ما حرج أو تردد. هذا بالإضافة إلى أنها تجنب الباحث خطأ التحيز الذى قد يظهر فى طريقة العمل الحقلى والاتصال المباشر بمفردات العينة أو المجتمع. وعلى الرغم من ذلك يعاب على طريقة المراسلة بالبريد أن نسبة الإجابات تكون عادة قليلة، وبصفة خاصة إذا كانت الاستمارة الإحصائية تحتوى على عدد كبير من الأسئلة فإن ذلك يكون سبباً فى إهمال المبحوثين للاستمارات وعدم استيفائها وإعادتها للباحث. كما تتضح صعوبة هذه الطريقة فى حالة إذا كان عدد من مفردات العينة يجهلون القراءة والكتابة. هذا بالإضافة إلى أن هذه الطريقة تحتاج إلى دقة بالغة فى وضع الأسئلة، إذ أنه قد ينتج عن عدم فهم المبحوثين بعض الأسئلة وقوعهم فى أخطاء تؤثر على دقة النتائج مثل خطأ التحيز فى الإدلاء بالمعلومات لانعدام الرقابة على الإجابات ولاعتقاد المبحوثين بعدم جدية أو ضرورة الدراسة.

ب- الاتصال التليفونى: تصلح طريقة الاتصال التليفونى، كطريقة من طرق جمع البيانات، فى الدراسات المحدودة التى يلعب فيها عامل الوقت دوراً مؤثراً، والتى يضمن فيها الباحث وجود أجهزة تليفونية لدى المبحوثين الذين تتكون منهم العينة أو المجتمع. ويمكن بهذه الطريقة مخاطبة المبحوثين والحصول منهم على الإجابات للأسئلة الموجهة إليهم.

وعلى الرغم من أن الإتصال التليفونى يعد من أسهل الطرق لجمع البيانات إلا أنها أصعبها من حيث إمكانية الحصول على نسبة عالية من الإجابات إذا كانت الأسئلة طويلة وتحتاج لوقت طويل فى فهمها، لذا فلا بد أن تكون الأسئلة قصيرة وسريعة حتى لاتأخذ وقتاً طويلاً فى الإجابة عليها.

ثانياً: العمل الحقلى (الميدانى) Fieldwork

العمل الحقلى أو الميدانى كما عرفه Wooldige and East (1967) هو عبارة عن «الفحص القريب والتحليل فى الميدان لجزء من البلاد، بما فيه من ظواهر

طبيعية وبشرية، تكون سهلة الوصول، وموضجاً مظهر أو أكثر من مظاهر الاختلاف المكاني، وبذلك يتميز العمل الحقلي بأنه يضع الباحث وجهاً لوجه أمام مفردات ومتغيرات الظاهرة أو الظواهر المراد دراستها وتحليلها، كما أن الباحث في الميدان يستطيع أن يرى ويلمس الجوانب غير الواضحة عن الظاهرة وبالتالي يتأكد من صحة البيانات والمعلومات السابقة عنها. ولذا فإن نجاح الباحث في دراسته يتوقف إلى حد كبير على نوعية وكيفية العمل الحقلي الذي أجراه، وعلى الوقت والجهد الذي بذله، وعلى الزيارات التي قضها في منطقة البحث.

ويشمل العمل الحقلي طريقة: المقابلة الشخصية (الاتصال المباشر) للمبحوثين، أو الملاحظة الميدانية وقياس الظواهر في الطبيعة، أو المسح الميداني والزيارات للمزارع والمصانع. ويتوقف استخدام كل طريقة منها على خطة البحث ونوع الدراسة، كما أن لكل منها مزايا وعيوب نوضحها فيما يلي:

أ- المقابلة الشخصية Interviewing:

تعرف المقابلة أحياناً بطريقة الاتصال المباشر لجمع البيانات، إذ يتم فيها انتقال الباحث إلى المبحوثين (مفردات العينة) وذلك بغرض المواجهة الشخصية للحصول على المعلومات التي تحتاجها الدراسة. كما في حالة دراسة الخصائص الاجتماعية والثقافية لسكان أحد الأقسام الإدارية في محافظة ما.

وفي حالة دراسة مفردات مجتمع يتميز بأن عدد مفرداته كبير يستعين الباحث بمندوبين لجمع البيانات يشترط فيهم أن يكونوا مدربين تدريباً كافياً على العمل بهذه الطريقة، ويتصفون بالأمانة في تدوين البيانات.

ويمكن أن تتم المقابلة إما في أشكال محددة، أو في صورة غير محددة. فهناك المقابلة المحددة أو المقفولة Closed Interview، وهي المقابلة المقتنة أو المنهجية، التي تتخذ أسلوباً منظماً حيث تكون حالة وضع الأسئلة سابقة على المقابلة نفسها، وتوجه نفس الأسئلة لجميع المفردات بدون تغيير سواء في الأسلوب أو الصياغة. وهناك أيضاً المقابلة غير المحددة أو المفتوحة، وهي المقابلة غير المقتنة أو غير المنهجية،

التي تتميز بالأسئلة الحرة التي تتواتر بطريقة طبيعية تلقائية، أى لا تلتزم باستخدام صياغة الأسئلة وأسلوبها.

ومن أهم مزايا المقابلة الشخصية أنها تلائم كثيراً دراسة المناطق التي ترتفع نسبة الأمية بين سكانها، كما تتيح للباحث الحصول على معلومات أولية تقل فيها الأخطاء الصادرة من المبحوثين إلى درجة كبيرة. كذلك تعطى هذه الطريقة الفرصة لتوضيح الأسئلة الغامضة أو التي تبدو غير مفهومة للمبحوثين، هذا بالإضافة إلى أنه يمكن للباحث إضافة بعض الأسئلة التي يرى إضافتها، أو حذف البعض الآخر تبعاً لما تمليه ظروف المقابلة، كما أن الباحث يستطيع كشف أى تناقض يمكن أن يقع فيه المبحوثين.

أما عيوب طريقة المقابلة فمن أهمها احتمال تحيز الباحث، أو توجيهه لمفردات المبحوثين لوجهة نظر لا تخدم غرض البحث مما يؤثر على دقة النتائج، أو قد تتضمن هذه الطريقة بعض الغش إذ من المحتمل أن يقوم الباحث باستكمال بعض الإجابات بنفسه ليتسنى له إتمام عمله فى وقت قصير. وتحتاج طريقة المقابلة جهوداً مفسنية واعتمادات مالية كبيرة إذ أنها تحتاج فى بعض الأحيان إلى عدد كبير من المندوبين لجمع البيانات. كما أن هذه الطريقة لا تتماشى مع الدراسات التي تتصف بأنها تأخذ طابعاً خاصاً، بمعنى أن تكون أسئلتها مخرجة أو حساسة للأفراد الذين يجرى عليهم الباحث والاستقصاء والذين ربما يحجمون عن الإجابة على مثل هذا النوع من الأسئلة.

ب- الملاحظة الميدانية وقياس الظواهر فى الطبيعة:

تعرف الملاحظة الميدانية بأنها المشاهدة الدقيقة لظاهرة ما، مع الاستعانة بأيساليب البحث والدراسة لقياس وتسجيل كافة أوجه التغيرات Variations (مكانية أو زمنية) فى الظاهرة وفق خطة معينة تتلائم مع طبيعة تلك الظاهرة. وهى بذلك تستخدم فى الدراسات والأبحاث التي تهدف إلى اختبار الأسس والنظريات Basis and theoris التي تضبط ظاهرة أو مشكلة معينة. فعند دراسة ظاهرة أو مشكلة ما، مثلاً، فإننا نضع الفروض المناسبة لدراستها وحلها على أساس وضع خطة محددة

تشتمل على بعض التجارب العلمية أو القياسات الحقلية، باستعمال بعض الأجهزة لقياس وتسجيل المتغيرات المتعلقة بالفروض الموضوعية، ثم نقوم باختبار صحة الفرض واحداً بعد الآخر مع استبعاد الفرض الذى لا تثبت صحته وأهميته.

وعند إجراء عملية الملاحظة فإنه يجب على الباحث مراعاة الحرص والدقة في القياس وعدم التحيز. فمثلاً عند تحليل الاختلافات المكانية أو الزمنية في شكل شاطئ منطقة ساحلية فإنه يجب قياس وتسجيل المتغيرات التي تؤثر في هذه الاختلافات حتى يمكن الوقوف على الطريقة التي يتأثر بها الشاطئ، ولرسم وتحديد المناطق التي تحدث بها هذه الاختلافات أكثر من غيرها من المناطق. ومن المتغيرات التي يمكن قياسها لهذا الغرض: انحدار الشاطئ Beach Slope حجم الرواسب، خصائص الأمواج (طول الموجة، ارتفاعها، انحدارها وقوتها، زمن الموجة واتجاهها)، وخصائص التيارات الساحلية من حيث اتجاه التيار وقوته، والعوامل الجوية من ضغط جوى واتجاه الرياح وقوتها.

ويسود استخدام الملاحظة أو المشاهدة كأداة من أدوات البحث لجمع وتسجيل المعلومات في مجال الدراسات المعملية أو في التجارب الحقلية، وذلك بقصد تحقيق فرض معين، أو لمعرفة علاقة بين متغيرين، أو لإيضاح بعض النتائج التجريبية التي يكون معناها مازال غامضاً أو مبهماً. وأثناء الملاحظة يقوم الباحث بتسجيل الملاحظات أو المشاهدات في شكل قياسات معملية أو حقلية عن المتغيرات التي تحكم التجربة مثل رصد الحركة والوقت، كما في حالة رصد حركة المرور على أحد الطرق في فترات زمنية محددة. وتصدر الإشارة هنا إلى أن الملاحظة أو المشاهدة عن طريق التجربة تمكن الباحث من السيطرة على بعض المتغيرات المحيطة بالظاهرة قيد البحث والتي لا تدخل ضمن المتغيرات التي يراد دراستها حتى يمكن اختبار سلوك الظاهرة بدقة، والوقوف على حقيقة التأثير النسبي لمختلف المتغيرات. ونظراً لصعوبة التحكم في جميع المحددات والعوامل الجغرافية، وهي عوامل متشابكة، فإن الملاحظة أو المشاهدات عن طريق التجربة الميدانية في الدراسات الجغرافية لا تتم إلا إذا كان مجال الظاهرة محدوداً والمتغيرات قيد البحث عددها قليلاً.

جـ- الزيارات:

يشمل العمل الحقلى (الميدانى) أيضاً جمع البيانات المنشورة التى تفيد دراسة وتحليل الظاهرة قيد البحث كالتقارير والوثائق والاحصاءات من الشركات والمؤسسات أو الهيئات الخاصة والحكومية فى منطقة البحث. وكما يقول Wool-drige and East (1967) أن «العمل الحقلى لا يعد عملاً حقيقياً إلا إذا شمل الزيارات للمزارع والمصانع ومراكز الاحصاء». وما تجدر الإشارة إليه فى هذا الصدد أنه يجب مطابقة النشرات والتقارير التى يجمعها الباحث من مصادرها على الطبيعة للوثوق من سلامتها العلمية، وللتأكد من صحة ما تحتويه من بيانات، كما يجب التعرف على الطرق والأساليب التى جمعت بواسطتها هذه البيانات ويتم ذلك عن طريق مناقشة المختصين وذوى الخبرة فى الهيئات المسؤولة عن نشر البيانات. وإذا ما تعدل الحصول على البيانات المطلوبة أثناء القيام بالزيارات الميدانية فإنه لا بد أن يقوم الباحث بتصميم استمارة احصائية، أو ما يعرف «بالاستبيان» لاستكمال هذا النقض بنفسه عن طريق الاستفسار الشخصى وتدوين المعلومات عن كل أو بعض المتغيرات المطلوب دراستها.

من العجالة السابقة عن العمل الحقلى نلاحظ أن إجراءاته ووسائله تتنوع بتنوع الظروف والعوامل المتحركة فى الظاهرة قيد البحث، ولكن قد يستعين الباحث فى الميدان ببعض الوسائل التى تعينه فى الدراسة الميدانية بصفة عامة وفى الملاحظة بصفة خاصة وهى:

(١) تسجيل القياسات التى أجريت على المتغيرات المطلوب دراستها بواسطة الأجهزة والمعينات الخاصة، وتدوين المشاهدات الميدانية إما فى جداول وإما على أجهزة تسجيل إذا تيسر استعمالها. ولتدوين القياسات والمشاهدات مكانة خاصة فى الدراسات المنظمة الهادفة لأهميتها فى تكوين العلاقات بين المشاهدات والملاحظات المختلفة.

(٢) الخرائط: تعتبر الخريطة من أصلح الوسائل لمعرفة العلاقات المكانية وتفسير وتعليل كثير من غوامض الظاهرة أو الظواهر المدروسة. ونظراً لأنه يستحيل

على الباحث أن يزور كل مكان ويفحصه في الطبيعة، فلا بد له من الاعتماد على الخريطة لمعرفة الأماكن التي يصعب عليه رؤيتها في الطبيعة. وعموماً فإن الباحث عند قيامه بالزيارات أو ملاحظة وقياس متغيرات ظاهرة ما في الطبيعة يحتاج إلى عدة خرائط ضرورة عن الظروف الطبيعية لمنطقة الدراسة من أهمها الخريطة الكنتورية والطبوغرافية، والخريطة المناخية، وخريطة استخدام الأرض في الأغراض المختلفة، وخريطة السكان. ومما يدل على أن الخريطة هي أداة الجغرافى الأولى فى العمل الحقلى ما قاله H.R. Mill فى ذلك «إن ما لا يمكن إثباته على خريطة لا يمكن وصفه» (Wooldrige and East, 1967). وبناء على ذلك أصبحت الخرائط ذات أهمية خاصة فى مختلف فروع العلم والمعرفة.

(٣) الصور الجوية والفتوغرافية لظواهرات ومواقف معينة: وهما من الوسائل المعينة فى الدراسة الميدانية. فمن المعروف أن الباحث قد يرى للموقف من خلال طريقة تفكيره الخاصة وإهتماماته الشخصية بالدراسة التي يقوم بها، ولكن لا يتفق باحثان على وصف واحد لظاهرة معينة أو لموقف معين إذ أن كلا منهما يراه من وجهة نظره الخاصة.

وعموماً إذا كان منهج البحث هو الذى يحدد الطريقة المتبعة فيه، فإن الطريقة - بالتالى - هى التى تتحدد أداة أو وسيلة جمع البيانات الأكثر مناسبة. إلا أن هذا لا يعنى الاعتماد على أداة واحدة فقط من الأدوات السابق ذكرها لجميع البيانات، لأنه يمكن جمع المعلومات بأكثر من أداة تبعاً لما تعلية طبيعة الدراسة ونوعية المعلومات المطلوب الحصول عليها. ونظراً لأن هذه الأدوات غير مستقلة تماماً عن بعضها فإن الباحث يستطيع أن يحدد الأدوات التى سيستخدمها فى جمعه للبيانات والمعلومات الدقيقة التى يتعرض لها بالتحليل الإحصائى بحيث يحصل فى النهاية على نتائج يتخذ على أساسها القرارات فيما بعد.

الاستمارات الإحصائية:

بعد أن يتم اختيار طريقة جمع البيانات والمعلومات عن خصائص مفردات

المجتمع أو العينة، سواء بأسلوب الحصر الشامل أو المعاينة (العينات)، يلجأ الباحث إلى عمل استمارة إحصائية خاصة، تتضمن أسئلة محددة عن تلك الخصائص المراد معرفتها وقياسها، ولتكون مرشداً له في جمع بياناته ورسم إطاراً محدداً لها، هذا فضلاً عن استخدامها كأداة لتسجيل البيانات أو قناة تستقي المعلومات من خلالها. وعادة ما تستخدم الاستمارة الإحصائية في الدراسات التي تحتاج إلى جمع بيانات كثيرة قابلة للقياس ويمكن تسجيلها بانتظام.

وكما ذكرنا آنفاً أن خطوات تصميم البحث تبدأ بوضع إطار للبيانات التي يجب الحصول عليها لاستخدامها في حل مشكلة البحث، ثم تحديد مصادر هذه البيانات والوسائل التي ستتيح في الحصول على هذه البيانات. وفي المرحلة الأخيرة ذكرنا أنه يجب أن تختار وسائل جمع البيانات اختياراً سليماً يبنى على أساس مدى ملائمة كل وسيلة، من حيث مزاياها وعيوبها، لأهداف البحث، وعادة ما تكون الاستمارة الإحصائية أداة هامة من أدوات أو وسائل جمع البيانات وتستخدم بالإضافة إلى الأدوات أو الوسائل الأخرى مثل المقابلة أو الملاحظة.

- وهناك نوعين رئيسيين من الاستمارات الإحصائية هما: كشف البحث Schedule ، وصحيفة الاستبيان Questionnaire ، ولكل منهما مزاياه وعيوبه التي نوضحها فيما يلي:

١- كشف البحث: يطلق اسم كشف البحث على الاستمارة الإحصائية التي تضم مجموعة من الأسئلة التي تسأل وتدون بواسطة الباحث في مقابلة شخصية للبحوث (مفردة البحث) الذي وقع عليه الاختيار في عينة البحث. كما يضم كشف البحث عند استخدامه في الملاحظة أو الملاحظة الميدانية بيان بمتغيرات الظاهرة المطلوب قياسها وجمع المعلومات عنها. ويتيح هذا النوع من الاستمارات الإحصائية للباحث درجة عالية من المرونة عند تصميمها بإعطاء الفرصة في إضافة أو حذف ما يراه الباحث من أسئلة تبعاً لظروف المقابلة الشخصية، أو صرف النظر عن بعض العوامل التي قد لا يكون لها دوراً يذكر في تبين الظاهرة موضع الدراسة. إلا أنه من أخطر عيوب كشف البحث هو احتمال تحيز الباحث لوجهة نظر شخصية لا تخدم غرض البحث، مما يؤثر على دقة النتائج التي ينتهي إليها البحث.

٢- صحيفة الاستبيان: وهى عبارة عن الأداة التى تستخدم للحصول على البيانات عن طريق الإجابة على أسئلة تتعلق بالظاهرة قيد البحث والتى يجيب عليها المبحوث بنفسه، وهذه قد ترسل بالبريد أو تسلم باليد للمبحوث الذى يطلب منه فى كلتا الحالتين إعادتها للباحث بعد إستيفائها.

وتجدر الإشارة هنا إلى توضيح الفرق بين الاستبيان وصحيفة الاستبيان حتى لا يختلط الأمر بينهما. فالأولى عبارة عن وسيلة قائمة بذاتها لجمع المعلومات بطريقة سريعة عن موضوعات محددة ومن مجموعة كبيرة من المفردات (المبحوثين)، بينما تستخدم الثانية كأداة لهذه الوسيلة التى يكون هدفها الأساسى ترجمة البحث العلمى إلى أسئلة معينة. وبصفة عامة تتميز صحيفة الاستبيان بسهولة تنفيذها وتوفيرها للوقت والتكاليف المادية، وإتاحتها الفرصة للمبحوث فى التفكير والإجابة على الأسئلة الحرجة دون تردد، بالإضافة إلى أنها تجنب للباحث الوقوع فى خطأ التحيز لعدم إمكانية فرضه لرأى معين أو لوجهة نظر خاصة. إلا أن إمكانية وجود أخطاء ناجمة عن تحيز المبحوث نفسه فى إجابة الأسئلة يعتبر من أهم مثالب صحيفة الاستبيان، بالإضافة إلى أنها لاتصلح تماماً إذا كانت مجموعة المبحوثين فى العينة أو المجتمع تحتوى على عدد كبير يجهل القراءة والكتابة، أو إذا كانت البيانات المطلوبة كثيرة ووقت المبحوث ضيقاً مما يؤدى إلى تكاسل المبحوث فى استيفاء الاستمارة وإعادتها للباحث.

تصميم الاستمارة الاحصائية:

مهما كانت طبيعة البيانات المطلوب الحصول عليها أو الوسيلة المتبعة فى جميع هذه البيانات فإنه يجب على الباحث مراعاة بعض الشروط الهامة عند تصميمه للاستمارة الإحصائية، لأن التصميم الجيد والصياغة المتقنة لأسئلة الاستمارة يعد أحد العوامل الجوهرية فى انجاح العمل الحقلى بصفة خاصة والبحث الذى يقوم عليه بصفة عامة. وتحتاج عملية تصميم الاستمارة الإحصائية من الباحث المعرفة الكاملة والدراية التامة بأصول صياغة الأسئلة. ورغم أن الاستمارات تختلف فى تصميمها، إلا أن هناك قواعد وشروط يجب توافرها حتى

يأخذ تصميم الاستمارة دورة في انجاح البحث. هذه الشروط منها ما هو متعلق بشكل الاستمارة، ومنها ما هو متعلق بمضمونها من حيث نوعية الأسئلة وطريقة وضعها وصياغتها.

١- شكل الاستمارة: لاشك أن الإهتمام بشكل الاستمارة الإحصائية يعتبر من العوامل الرئيسية في عملية جمع البيانات الدقيقة غير المشكوك فيها، حيث يشجع الشكل الجيد المبحوثين على الاستجابة لمحتواها. ويتحدد شكل الاستمارة الجيد بعدة عوامل منها:

أ- جودة الاستمارة، من حيث نوع الورق المستخدم الذي يجب أن يكون من النوع الذي يتحمل الاستخدام الكثير من تدوين المعلومات.

ب- حجم الاستمارة، من حيث عدد صفحات الاستمارة التي يجب أن لا تكون قليلة على حساب الأماكن الحالية المخصصة للإجابة، أو لا تكون كثيرة حتى لا يكون ذلك سبباً في إرهاق المبحوثين في الإجابة على أسئلتها.

ج- ترتيب وتنظيم الأسئلة داخل الاستمارة، إذ أن التسلسل والترتيب في وضع الأسئلة (عن طريق إعطاء الأسئلة أرقاماً تدريجية، أو وضع الأسئلة في شكل مجموعات أو تقسيمات متجانسة تترايط فيما بينها ترابطاً منهجياً، يمكن معه حصر المطلوب، بحيث تبدأ من الأسئلة البسيطة إلى الأسئلة المركبة، أو من أسئلة عامة تتميز بالشمول، إلى أسئلة خاصة تتميز بالتركيز على أفكار دقيقة ومحددة). يعتبر من أهم الشروط التي يجب مراعاتها عند تصميم الاستمارة الإحصائية مهما كان نوعها، لأن ذلك يساعد على سهولة الإجابة، كما يعمل على تسهيل عملية التحليل والدراسة بعد ذلك.

وعموماً يجب أن يظهر عنوان البحث بوضوح في صدر الاستمارة، وكذلك اسم الهيئة أو الجهة المشرفة على الدراسة، بالإضافة إلى ما يشير إلى سرية استخدام بيانات الاستمارة إلا لغرض البحث فقط، مع وضع بعض التعليمات المختصرة والمبسطة لتوضيح أهداف الدراسة إن أمكن ذلك. ونظراً لأن معظم التحليلات

الاحصائية تقوم بها فى الوقت الحاضر أجهزة الحاسب الآلى فمن المستحسن أن تتضمن الاستمارة رموزاً Codes حتى تسهل مهمة نقلها وتفرغها على البطاقات الخاصة بالحاسب الآلى.

٢- مضمون الاستمارة: يقصد بمضمون الاستمارة هو كيفية صياغة الأسئلة والتي تعد ذات أهمية بالغة فى الحصول على إجابات صحيحة وبالتالي على معلومات دقيقة. وكلما كانت الأسئلة، أو التعبير عما هو مطلوب واضح دون ما صعوبة أو تعقيد لفظى أو سوء فهم كلما سهلت مهمة الباحث والمبحث فى نفس الوقت. وبصفة عامة فإنه يمكن تحقيق ذلك بأن تكون الأسئلة على شكل حوار طبيعى تلقائى. أى ليس المقصود بها أن نتوصل إلى إجابات معينة، مع تجنب الأسئلة الطويلة التي تزيد من احتمالات سوء الفهم - كما يجب أن تكون الأسئلة محددة ودقيقة حتى نحصل على معلومات صحيحة، أى يجب أن يعطى كل سؤال فكرة واحدة واضحة عما نطلب السؤال عنه. فمثلاً يبدو السؤال: أين كان ميلادك؟ غامضاً، والأفضل منه يكون السؤال: فى أى قرية أو مدينة كان ميلادك؟ وهو يبدو أكثر تحديداً ووضوحاً من السؤال الأول. كما يجب أن تكون الأسئلة بعيدة تماماً عن الأسئلة الحرجة ذات الحساسية البالغة. ويستطيع الباحث التحايل على ذلك بصياغة أسئلة غير مباشرة، فمثلاً يمكن التعرف على مقدرة ودخل العامل بطرح الأسئلة التي تستفسر عن طبيعة العمل الذي يقوم به العامل. على أنه يجب أن يكون الباحث لبقاً وذكياً عند وضع الأسئلة حتى لا يضع أسئلة تروحي بإجابات معينة، أو أسئلة افتراضية تكون الإجابة عليها غير مفيدة. فمثلاً يمكن طرح السؤال: ما مقدار الأجر الإضافى الذي ترغب أن تحصل عليه شهرياً حتى يتحسن مستوى معيشتك؟ بدلاً من السؤال: هل تكون راضياً لو ارتفع مرتبك الشهرى إلى ٦٠ جنيه؟. كذلك يجب وضع تفسيرات محددة للمصطلحات التي تكون مجالاً للشك من حيث الفهم، وتوضيحات دقيقة للتعريفات المستخدمة مثل تعريف الأسرة أو الدخل. كما يجب أن تصاغ الأسئلة إما لتوضيح الآراء أو

الاتجاهات Attitudes أو لتوضيح الحقائق مثل السن والمهنة أو الملكية الزراعية أو العقارية.

وقبل إتمام صياغة الاستمارة الاحصائية، ينبغي على الباحث أن يتفهم طبيعة المبحوثين موضع الدراسة وذلك عن طريق تصميم استمارة استطلاعية - Pilot Questionnaire توزع على عينة ذات عدد محدود من الأفراد ليست لهم علاقة بالبحث ليجبروا على أسئلتها، ومن طريقة الإجابة في الاستمارة الاستطلاعية يمكن التعرف على الأسئلة التي يمكن أن تكون غامضة أو غير مفهومة لتعاد صياغتها بعد توضيحها. كما تجدر الإشارة في هذا الصدد إلى أنه يجب على الباحث أن يضع بعض الأسئلة للمراجعة Checking Questions للتأكد من صحة الإجابات، خصوصاً عند وجود تعارض في الإجابات على هذا النوع من الأسئلة وإجابات الأسئلة الخاصة في الاستمارة والتي تحمل نفس الإجابة أو عكسها. فمثلاً السؤال: هل تحب عملك؟ يتعارض مع السؤال: هل تتغيب كثيراً عن العمل؟. فإذا كانت الإجابة على السؤال الأول بالإيجاب وعلى الثاني بالنفي فإن ذلك يؤكد أن حب العمل لا يؤدي إلى التغيب كثيراً عن العمل.

وفيما يلي مثال لاستمارة إحصائية عن دراسة العمران في إحدى قرى منطقة زراعية مستصلحة حديثاً:

جامعة الإسكندرية
كلية الآداب
قسم الجغرافيا

استمارة بحث رقم (١)

دراسة العمران

(هذه الاستمارة سرية للغاية ولا تستخدم بياناتها إلا فى الأغراض العلمية)

أولاً: الحالة الاجتماعية للوافدين:

(١) عدد الأسر فى المسكن ()

(٢) (٣) (٤) (٥) (٦) (٧) (٨)

الاسم النوع البن المهنة الحالة الاجتماعية المؤهل محل الميلاد الدخول الشهرى

-١

-٢

-٣

-٤

-٥

-٦

-٧

-٨

-٩

-١٠

-١١

-١٢

-١٣

-١٤

-١٥

(٩) الجهة الوافدة منها:

(١٠) سنة القدوم إلى القرية:

(١١) المهنة عند القدوم:

(١٢) هل تسافر إلى موطنك الأصلي:

ثانياً: السكن:

ب- ملكية السكن: (١) ملك خاص (٢) ملك أهالي (٣) ملك حكومي

(٤) الإيجار الشهري (٥) المساحة التي يشغلها المبنى

ج- ارتفاع المبنى: (١) دور واحد (٢) دورين (٣) أكثر من دورين (٤) عدد الغرف

(٥) هل توجد حظيرة الحيوان بداخله أم توجد خارجه

د- المرافق الصحية: (١) داخل المسكن.

(٢) خارج المسكن.

(٣) هل المطبخ حجرة مستقلة أو لا يوجد

(٤) هل الحمام مستقل مع المراض لا يوجد

(٥) هل المراض مستقل مع الحمام لا يوجد

(٦) هل المنزل به صرف صحي - نعم لا

هـ- مادة البناء: (١) طوب أحمر (٢) حجر جيري (٣) مواد أخرى

(٤) هل الأرضية بلاط أسمنت خشب تراب مواد أخرى

(١) هل المبنى له توصيلة مياه أو بدون توصيلة

و- مياه الشرب: (٢) ما مصدر المياه

ز- الكهرباء: (١) هل المبنى به توصيلة كهرباء أو بدون كهرباء

الفصل الثاني

تصنيف وجدولة البيانات

عندما يتجمع لدى الباحث كمية كبيرة من الحقائق أو البيانات الإحصائية في صورة غير منتظمة والتي تم جمعها سواء عن طريق الحصر الشامل Survey كالتعداد السكاني Population Census، أو عن طريق العينات Samples، فإنه يتعذر عليه إستيعاب أو استخلاص النتائج من هذه البيانات قبل تصنيفها وتنظيمها وتلخيصها في صورة جدولية. ويتوقف تصنيف وجدولة البيانات على طبيعتها والغرض من البحث الذي يسعى الباحث إلى تحقيقه.

ولا توجد طريقة موحدة يمكن الاعتماد عليها في جدولة البيانات، ولكن هناك قواعد عامة يجب مراعاتها عند تصميم جداول البيانات الإحصائية، وهي:

- ١- أن يكون عنوان الجدول محددا لما يتضمنه من بيانات.
- ٢- أن ترتب البيانات بالجدول حسب أهميتها أو تسلسلها الزمني، مع توضيح وحدات القياس التي جمعت على أساسها.
- ٣- أن يوضح المصدر الذي اعتمد عليه في تكوين بيانات الجدول.

تجميع البيانات:

بعد اكتمال عملية البيانات يتم تجميع المعلومات من الاستمارات الإحصائية ككشف البحث أو استمارة الاستبيان، ووضعها في استمارة خاصة تعرف باستمارة التجميع أو التفريع Master Sheet، حيث يعطى فيها كل متغير رمزا

خاصاً به Coding. وعادة ما نحرر استمارة التجميع على ورق المربعات حتى تسهل عملية التفريغ، ويقوم الباحث بتقسيمها إلى محورين: الأول يشمل أنواع المتغيرات التي تندرج تحتها جميع البيانات التي تم جمعها. ويشمل المحور الثاني أرقام أو رموز الاستمارات الإحصائية التي جمع فيها البحث بيانات بحثه. فمثلاً إذا كانت أسئلة الاستمارة الإحصائية تتضمن بيانات عن التركيب العمري والجنس وحجم الأسرة ودرجة التعليم ممثلة بعدد سنوات الدراسة والمستوى الاقتصادي ممثلاً بالدخل السنوي للمبحوثين، فإن الباحث يقوم بإعطاء كل متغير من هذه المتغيرات رقماً خاصاً به، كأن يعطى الرقم (١) ليمثل التركيب العمري، والرقم (٢) ليمثل الجنس والرقم (٣) ليمثل حجم الأسرة ... وهكذا. ويمكن أن ترتب مثل هذه البيانات كما في الجدول التالي:

جدول رقم (٢-١)

طريقة تجميع المعلومات من الاستمارة الإحصائية

رقم الاستمارة	(١) التركيب العمري			(٢) الجنس		(٣) حجم الأسرة	(٤) درجة التعلم		(٥) الدخل السنوي جنية
	أقل من ١٥	من ١٥ إلى ٣٠	أكثر من ٣٠	ذكر	أنثى		متعلم	غير متعلم	
١			٥٣	ذكر		٤	١٦		١٤٤٠
٢		٢٣			أنثى	٢	١٢		٦٦٠
٣			٤٢	ذكر		٣	١٨		١٠٨٠
٤			٦١	ذكر		٥	٩		٨٤٠
٥			٣٦	ذكر		٤	٢٠		١٨٠٠
٦		٢٥		ذكر		٣	٦		٤٢٠

وهكذا فإن الاستمارة رقم (١) تدل على المبحوث الأول الذي يكون عمره ٥٢ سنة وحجم أسرته ٤ أفراد وهو متعلم قضى في التعليم ١٦ سنة ودخله السنوي

١٤٤٠ جنيهاً. ويطلق على هذا النوع من الاستثمار اسم «جدول التفريغ» حيث يتم تفريغ المعلومات به بعد كتابة عنوانه ووحداته ومصدره.

تصنيف البيانات:

يعتبر تصنيف Classification أو ترتيب البيانات وتقسيمها تقسيماً سهلاً لإدراك ما بينها من علاقات ويوضح صفاتها ومميزاتها، أساساً لأى أسلوب من أساليب التحليل العلمى ويتم تصنيف المعلومات والبيانات على أسس مختلفة يحددها شكل البحث من جهة والخصائص المميزة للبيانات من جهة أخرى. ويمكن الإشارة هنا إلى خمسة أنواع من التصنيفات ترتب على أساسها البيانات، وهى:

١- التصنيف الأبجدي، ويتم فيه ترتيب مفردات الظاهرة، كترتيب الدول مثلاً، حسب أبجديتها، فمثلاً تسبق دولة أندونيسيا دولة بورما ثم تأتى جامايكا والجزائر وهكذا.

٢- التصنيف الزمني، وفيه ترتب المعلومات حسب الفترات الزمنية (سنوات)، فإذا كان المقصود دراسة تطور قيمة المنتجات الصناعية فى مصر فى الفترة من ١٩٥٧ إلى ١٩٦١، فترتب المعلومات متتالية حسب السنوات. ويوضح ذلك الجدول الآتى:

جدول رقم (٢-٢)

قيمة المنتجات الصناعية فى جمهورية مصر العربية

فى السنوات ١٩٥٧ - ١٩٦١

السنة	قيمة المنتجات (بالآف الجنيهات)
١٩٥٧	٢٩٢٢٤٩
١٩٥٨	٤١٨٥٩٧
١٩٥٩	٤٤٨٧٦٦
١٩٦٠	٤٩٢٧٤٥
١٩٦١	٦٤٧٩٩٢

(المصدر: الإدارة المركزية للإحصاء (مصلحة الإحصاء والتعداد) إحصاء الإنتاج للصناعى)

٣- التصنيف الجغرافى، أى تقسيم العالم إلى وحدات جغرافية كبرى (قارات ومحيطات)، أو تقسيم الدول إلى وحدات جغرافية (أو مقاطعات) والجدول التالى يوضح هذا النوع من التصنيف.

جدول رقم (٢-٣)
مساحات قارات ومحيطات العالم

القارة	المساحة مليون كيلومتر مربع	المحيط	المساحة م.ك
أفريقيا	٣٠,٣	الهادى	١٨٣,٤
آسيا	٢٦,٩	الاطلنطى	١٠٦,٧
أوروبا والاتحاد السوفيتى	٢٥,٤	الهندي	٧٣,٨
أمريكا الشمالية	٢٤,٣	القطبى الجنوبي	١٩,٧
أمريكا الجنوبية	١٧,٩	القطبى الشمالى	١٢,٤
استراليا ونيوزيلندا	٨,٥		
المجموع	١٣٣,٣	المجموع	٣٨٦,٠

٤- التصنيف الكيفى أو النوعى Qualitative وفيه يتم تقسيم المفردات حسب خصائصها، كأن يصنف الأفراد حسب الحالة الزوجية إلى متزوج، مطلق، أرمل، ولم يتزوج أبداً (أعزب). أو حسب الحالة التعليمية إلى أمى، يقرأ ويكتب ومتعلم. أو تصنف الأقاليم الجغرافية حسب درجات الحرارة إلى أقاليم حارة، أقاليم معتدلة وأقاليم باردة، أو حسب شدة سقوط المطر إلى أقاليم جافة وأخرى رطبة، أو حسب الارتفاع إلى أقاليم سهلية وأخرى جبلية.

٥- التصنيف الكمي Quantitative، أى إعطاء قيم أو رتب رقمية للخصائص هو وضع الدراسة وترتيبها تصاعدياً أو تنازلياً. ويتم التصنيف على أساس «متغير أحصائى» قابل للقياس الكمي. ثم نوضع بيانات المتغيرات بعد تبويبها فى مجموعات متساوية (أو متقاربة) فى جدول يسمى بالجدول التكرارى. Fre- quency Table، كما فى الجدول التالى:

جدول رقم (٢-٤)

توزيع ارتفاعات قمم التلال المحيطة بشمال ويلز
(أخذت هذه البيانات من خريطة أطلس بمقياس ١: ٢٥,٠٠٠)

عدد التلال	فئات الارتفاع (متر)
٦٢٠	أقل من ٢٠٠
٦٠٠	٢٠٠ - ٣٩٩,٩
٤٣٠	٤٠٠ - ٥٩٩,٩
٤٠٠	٦٠٠ - ٧٩٩,٩
٤٥٠	٨٠٠ - ٩٩٩,٩
٢٥٠	١٠٠٠ - ١١٩٩,٩
٢٥٠	١٢٠٠ - ١٣٩٩,٩
٢٠٠	١٤٠٠ - ١٥٩٩,٩
٦٠١	أكثر من ١٦٠٠

ويجدر الإشارة إلى أنه قد يشتمل التصنيف على أكثر من نوع واحد من الأنواع السابقة. فقد يكون التصنيف زمنياً وجغرافياً في نفس الوقت، كأن يتضمن الجدول بيانات ظاهرة ما لعدد من السنوات موزعة حسب الوحدات الجغرافية للمنظمة موضع الدراسة. وقد يكون التصنيف نوعياً وجغرافياً في نفس الوقت أيضاً، كأن يتضمن الجدول بيانات خصائص الظاهرة أو أقسامها الفرعية وذلك بالنسبة لكل وحدة جغرافية للمنظمة. وقد يكون التصنيف نوعياً وكمياً في ذات الوقت إذا تضمن الجدول بيانات عن حجم أو كمية الظاهرة وذلك بالنسبة لكل قسم أو صفة من صفات نفس الظاهرة.

الجدولة اليدوية للبيانات الإحصائية:

يقصد بجدولة البيانات الإحصائية أو ما يعرف «تتويب البيانات»، وضعها أو

عرضها في أشكال جدولية مختلفة لغرض التبسيط وتلخيص خصائص المعلومات، وبطريقة تسهل دراستها وتحليلها باستخدام المقاييس الإحصائية والأساليب الكمية المختلفة للوصول إلى نتائج واتخاذ القرارات التي يمكن تعميمها على المجتمع الإحصائي. وتختلف طرق العرض الجدولي اليدوية (غير الآلية) باختلاف الأسلوب المستخدم، كما تتنوع الجداول الإحصائية باختلاف طبيعة البيانات المراد جدولتها. ويمكن تقسيم الجداول الإحصائية إلى نوعين رئيسيين هما: الجداول العادية، والجداول التكرارية.

الجداول العادية:

تقسم الجداول العادية عموماً إلى ثلاثة أنواع:

١- الجداول البسيطة، وهي التي تتمثل فيها الظاهرة وخصائصها الكمية الممثلة بالقيم والأرقام، كما هو في الجدول التالي:

جدول رقم (٢-٥)

توزيع ٢٠ أسرة حسب الحجم

حجم الأسرة وفرد	عدد الأسرة
٢	٢
٣	٢
٤	٦
٥	٥
٦	٢
٧	٢
٨	١
الجملة	٢٠

٢- الجداول المركبة، وهي التي تتمثل فيها الظاهرة وخصائصها الممثلة بالأرقام داخل عدة أعمدة الجدول كأن نحدد عدد سكان الحضر والريف في الوحدات الجغرافية الكبرى لجمهورية مصر العربية.

جدول رقم (٢-٦)
التوزيع الجغرافى لسكان كل من الريف
والحضر فى مصر (تعداد ١٩٦٠)

المحافظات	عدد السكان بالآلاف		المجموع
	حضر	ريف	
محافظات الحضر	٥٥٨٢	-	٥٨٨٢
الوجه البحرى	٢٠٥٠	٨٧٨٨	١٠٨٣٨
الوجه القبلى	١٨٩٨	٧٣٣١	٩٢٢٩

٣- الجداول المزدوجة، وهو ذلك النوع من الجداول الذى يبين توزيع البيانات حسب صفتين فى نفس الوقت. وفيه تمثل الأعمدة تقسيمات أحد الصفتين، بينما تمثل الصفوف تقسيمات الصفة الأخرى. ومن أوضاع الأمثلة لهذا النوع جداول Input - Output Tables مثل الجداول الخاصة بالميزانيات المصرفية (الواردات - المصروفات) أو جداول الإنتاج والإستهلاك أو جداول توزيع الدخل على فئات عمر السكان. كما يظهر فى الجدول التالى:

جدول رقم (٢-٧)
فئة الدخل (بالجنيهات) والأعمار بالسنين

الدخل / السن	-٣	-٥	-٧	-٩	-١١	المجموع
-٢٠	٤	٢				٦
-٣٠		٤	٧			١١
-٤٠			٥	٣		٨
-٥٠			٠	٣	٢	٥
المجموع	٤	٦	١٢	٦	٢	٣٠

الجداول التكرارية: Frequency Tables

عند تلخيص كمية كبيرة من البيانات ذات الصفة الكمية المتغيرة فإنه من المفيد ترتيبها أو توزيعها على فئات وتحديد عدد المفردات التي تنتمي لكل فئة. والجدول المنظم على صورة فئات يقابل كل فئة تكرار معين يسمى جدول التوزيع التكرارى كما فى الجدول التالى:

جدول رقم (٢-٨)
توزيع تكرارى لأوزان ١٠٠ طالب

الأوزان (كيلو جرام)	عدد الطلبة (التكرار)
٦٠ - ٦٢	٥
٦٣ - ٦٥	١٨
٦٦ - ٦٨	٤٢
٦٩ - ٧١	٢٧
٧٢ - ٧٤	٨
المجموع	١٠٠

وتتلخص القواعد العامة لتكوين جدول التوزيع التكرارى فى الخطوات التالية:

١- تحديد أكبر وأقل قيمة فى البيانات ومنها توجد المدى (الفرق بين أكبر رقم وأقل رقم).

٢- تقسيم المدى إلى عدد مناسب من الفئات المتساوية الطول. وإذا لم يكن ذلك ممكناً تستخدم فئات ذات أطوال مختلفة أو فئات مفتوحة. ويؤخذ عدد الفئات عادة بين ٥، ٢٠ فئة حسب البيانات المتاحة. وتختار الفئات أيضاً بحيث يتفق مركز الفئة مع المشاهدات الفعلية. وهذا يؤدي إلى التقليل من أخطاء التجميع عند إجراء مزيد من المعالجة الإحصائية.

٣- تحديد عدد المشاهدات (أو التكرارات) التي تقع في كل فترة فئة وأحسن طريقة لأداء ذلك هو استخدام كشف الحزم أو العلامات.

وتسمى البيانات المنظمة والملخصة في جداول التوزيع التكرارى بالبيانات المجمعة. وعلى الرغم من أن عملية التجميع تؤدي بشكل عام إلى ضياع كثير من التفاصيل الأصلية للبيانات فإن الفائدة الهامة منها هي الصورة العامة التي يمكن الحصول عليها والتي تسهل تحليل واستخلاص النتائج والحقائق من البيانات.

اختيار الفئات:

لا توجد قاعدة ثابتة لتحديد أطوال وعدد الفئات، حيث أن ذلك يتوقف على طبيعة البيانات وعلى كمية الاختلاف بين المفردات وعلى الدقة المطلوبة في العمليات الحسابية. ولكن عند تحديد فترة (طول) الفئة فإنه يحسن الأخذ في الاعتبار العدد الكلى للقيم. فكل فترة الفئة يؤدي إلى الحصول على عدد من الفئات أقل مما لو كانت فترة الفئة أصغر. كما أنه عند تحديد عدد الفئات ينبغي اختيار عدد قليل من الفئات يؤدي إلى تسهيل العمليات الحسابية. ولتحديد عدد الفئات تستخدم إحدى الطرق الآتية:

$$١- \text{معادلة Yule} = \text{عدد الفئات} = \sqrt[5]{N}$$

حيث N تمثل عدد مفردات البيانات. وتصلح هذه المعادلة عندما يكون حجم العينة أقل من ١٠٠٠ مفردة.

٢- معادلة استرجس Sturges : عدد الفئات - ١ + (٣,٣٢٢ + $\log N$) حيث $\log N$ تمثل لوغاريتم عدد القيم. وحيث أن هذه الصيغة تعطي عدد ١٠ كسور عشرية، فيجوز هنا تقريب الناتج للحصول على مقدار صحيح لعدد الفئات. وتصلح هذه المعادلة عندما تزيد مفردات البيانات عن ١٠٠٠ مفردة.

وتلعب الخبرة في مجال تحديد عدد الفئات دوراً كبيراً. فلا يجب تقليل عدد الفئات حتى لا يؤدي ذلك إلى فقد بعض التفاصيل الموجودة في البيانات الخام، كذلك لا يجب زيادة عدد الفئات بحيث يصعب معها عملية الدراسة والتحليل

بعد ذلك. على أنه تجدر الإشارة إلى أن أنسب الجداول التكرارية هو الذى يحتوى على عدد من الفئات يتراوح بين ٨ إلى ١٢ فئة. بل يجب أن لا يتضمن الجدول على أقل من ٦ فئات ولا يزيد عن ٢٠ فئة.

فترة وحدود الفئات: Class (Cell) Interval and Limits

يعرف الرمز الذى يعبر عن الفئة ٦٠-٦٢ فى الجدول رقم (٢-٨) بفئة الفئة، كما يطلق على الرقمين ٦٠، ٦٢ حدود الفئة. ويسمى الرقم الأصغر (٦٠) بالحد الأدنى للفئة، والرقم الأكبر (٦٢) بالحد الأعلى للفئة. وفى معظم التوزيعات التكرارية تكون فترات الفئات متساوية. ويمكن تحديد فترة الفئة بالصورة الآتية:

$$\text{فترة (طول) الفئة} = \frac{\text{(الفرق بين أكبر وأصغر قيمة فى البيانات)}}{\text{عدد الفئات}}$$

وحيث أن هذه الصيغة تعطى طول فترة به كسور عشرية، فمن المستحسن تقريب الناتج للحصول على مقدار (أو رقم) صحيح لفترة الفئة.

ويمكن أيضاً تقدير فترة الفئة مباشرة بدون حساب عدد الفئات وذلك عن طريق العلاقة بين فترة الفئة والانحراف المعياري للبيانات. فمن المعروف من هذه العلاقة أن فترة الفئة والانحراف المعياري للبيانات. فمن المعروف من هذه العلاقة أن فترة الفئة عادة تتراوح عن $\frac{1}{4}$ و $\frac{1}{2}$ الانحراف المعياري للقيم، أو بمعنى آخر لا تقل فترة الفئة عن $\frac{1}{4}$ عن الانحراف المعياري ولا تزيد عن $\frac{1}{2}$ الانحراف المعياري. ولقد تم حساب الانحراف المعياري للبيانات المستمرة بمعرفة كل من عدد المفردات فى العينة والمدى بينها، ووضع فى جداول خاصة يطلق عليها اسم جداول Tippet كما فى الجدول التالى:

جدول رقم (٢-٩)

قيمة الانحراف المعياري للأحجام المختلفة من العينات

حجم العينة	٥٠	٧٥	١٠٠	١٥٠
الانحراف المعياري	٠,٢٢٢	٠,٢٠٨	٠,١٩٩	٠,١٨٩

حجم العينة	٢٠٠	٣٠٠	٥٠٠	٧٠٠
الانحراف المعياري	٠,١٨٢	٠,١٧٢	٠,١٦٤	٠,١٥٩

ى = المدى

فعلى سبيل المثال إذا كان العينة ١٠٠ والمدى بين المفردات ٤٦ فإن فترة الفئة يمكن تحديدها كما يلي:

$$\text{الانحراف المعياري للبيانات} = ٠,١٩٩ \times ٤٦ = ٩,١٥٤$$

∴ فترة الفئة تتراوح بين (٩,١٥٤) ، (٩,١٥٤) أى ما بين ٢ ، ٥ تقريباً.

وبعد اختيار عدد الفئات وتحديد فتراتهما يلى ذلك تحديد بداية ونهاية كل فئة وهو ما يسمى بحدود الفئات Class Boundaries. وتوضح حدود الفئات بحيث يكون الحد الأدنى للفئة الأولى أقل قليلاً من أقل قيمة فى البيانات، أما الحد الأعلى للفئة فيجب أن ينتهى قبل بداية الفئة التالية وهكذا بالنسبة لبقية الفئات حتى لا تتداخل حدود الفئات ويصبح من الصعب توزيع البيانات.

ومن الناحية النظرية تتضمن فترة الفئة كل القياسات بين حدى الفئة إذا كان المتغير موضع الدراسة من نوع المتغيرات غير المستمرة (الوثابة) التى لا تأخذ فيها كسرية كما فى الجدول رقم (٢-٨). فيما أن الأوزان سجلت لأقرب كيلو جرام فإن فترة الفئة ٦٠ - ٦٢ تشتمل على كل القياسات من ٥٩,٥ إلى ٦٢,٥، وهذه الأرقام تسمى بالحدود الحقيقية للفئة. والرقم الأصغر ٥٩,٥ هو الحد الأدنى الحقيقى للفئة التالية والقسمه على ٢. ويجب أن تكتب الفئات بحيث لا تتطابق

مع أحد القيم الفعلية. فإذا كانت الفئات المختلفة بالعمود الأول في الجدول رقم (٨-٢) مكتوبة بالصور ٦٠-٦٢، ٦٢-٦٥، ٦٥-٦٨، وكان لدينا القيمة ٦٥ فإنه يكون من الصعب تقرير ما إذا كانت تنتمي إلى الفئة ٦٢-٦٥ أو الفئة ٦٥-٦٨.

وفي حالة إذا كان المتغير قيد البحث من المتغيرات المستمرة (المتصلة) فيكون من الخطأ كتابة حدود الفئات بالصورة التي تكتب بها في حالة فئات المتغير غير المستمر. ويمكن كتابة حدود الفئة السابقة لها دون أن يحدث تداخل أو تترك ثغرات بين الفئات. وتكتب الفئات كمايلي: ٦٠ إلى أقل من ٦٣، ٦٣ إلى أقل من ٦٥، ٦٥ من ٦٨.... وهكذا. وللاختصار تكتب الحدود الدنيا للفئات وتترك حدودها العليا، إلا أنه في هذه الحالة يجب تحديد نهاية الفئة الأخيرة كمايلي: ٦٠، -٦٢، -٦٥، ٧١، ٧٤ إلى أقل من ٧٤.

مركز الفئة : Class mid-point

مركز الفئة هو منتصف فترة الفئة ونحصل عليه بجمع الحد الأدنى والحد الأعلى للفئة وقسمة المجموع على اثنين. فمركز الفئة ٦٠-٦٢ هو $\frac{٦٠ + ٦٢}{٢}$.

ويسمى مركز الفئة أيضاً بمنتصف الفئة. وعند التحليل الاحصائي للبيانات الموزعة تكراريا داخل فئات فإنه يفترض أن جميع التكرارات الموجودة داخل فترة فئة تأخذ قيما تتطابق مع مركز الفئة. وعلى ذلك فإن جميع الأوزان داخل الفئة ٦٠-٦٢ في الجدول رقم (٨-٢) تعتبر كما لو أنها ٦١ كيلو جرام.

تكوين الجدول التكرارى:

بعد اختبار عدد الفئات وتحديد فتراتها وحدودها نبدأ في تكوين جدول يشتمل على عدد من الصفوف مساو لعدد الفئات، ثم توزع البيانات الخام على الفئات المختلفة حسب قيمتها وحصر عدد المفردات الموجودة داخل فترة كل فئة باحدى الطرق الآتية:

١- طريقة المنظومة Array، وفيها يتم ترتيب وتنظيم البيانات إما تنازليا أو تصاعديا

حسب قيمتها مما يسهل تحديد عدد المفردات داخل كل فئة. وعلى الرغم من ذلك فإن هذه الطريقة لا يمكن تطبيقها على البيانات ذات المفردات الكثيرة إذ تصبح غير عملية في هذا الشأن.

٢- طريقة الحزم أو العلامات Tally marks، وفيها يتم توزيع مفردات البيانات مباشرة على الفئات وذلك بوضع (/) أمام الفئة المناسبة ثم نحصر عدد هذه العلامات عند نهاية تفريغ جميع مفردات البيانات. ويلاحظ أثناء التوزيع أننا نضع كل أربع علامات بجوار بعضها أما العلامة الخامسة فتوضع فوق الأربع علامات السابقة وبذلك تكون حزمة (||||) (أنظر الجدول رقم ٢-١١). والغرض من ذلك هو تسهيل عد العلامات داخل كل فئة. ويطلق على عدد العلامات أمام كل فئة إسم «تكرار الفئة Frequency» ويرمز له بالرمز (ك). وعند العرض النهائي للتوزيع التكراري يحذف عادة كشف الحزم (أو العلامات) من جدول ترتيب البيانات.

وسوف نوضح خطوات عمل جدول التوزيع التكراري للبيانات التالية والتي تمثل أطوال مائة راقد نهري (بالمتر) لأحد أحواض التصريف النهري.

جدول رقم (٢-١٠)

أطوال مائة راقد نهري (بالمتر)

٧٥٠	١١٢٠	١٠٠٠	١١٦٠	٩٩٠	١١١٠	٨٥٠	٨٢٠	١٠٨٠	٨٥٠
٩٤٠	٩١٠	١١٨٠	١٠٣٠	١٠٢٠	١٣٣٠	٩٨٠	١٠٦٠	٩٢٠	١٠٢٠
١١٥٠	١٠٩٠	١٠٠٠	٥٧٠	١٠٨٠	٧٧٠	٩٤٠	١٢١٠	١٠٠٠	١٠٧٠
١٠٤٠	٦٧٠	١١١٠	٨٨٠	٨٧٠	٩٧٠	١٠٢٠	٩٨٠	١٠١٠	٨٨٠
٩٠٠	٩٣٠	٨٥٠	١٠٧٠	٨٠٠	١٠٦٠	١٢٠٠	٩١٠	١٠١٠	١٠٣٠
١٠٩٠	١٠٠٠	١٢٧٠	١٠٧٠	١١٢٠	٩٨٠	٨٣	٩٨٠	٨٩٠	١٠٦٠
٧٩٠	١١٧٠	٨٥٠	٩٤٠	١١٩٠	٩٣٠	١٠٠٠	٩٠٠	١٠٢٠	٨٧٠
٩٥٠	١٠٩٠	١٤٢٠	٩٤٠	٩٣٠	٧٢٠	٩٨٠	١٠٥٠	١٢٢٠	١٠٤٠
١٠٤٠	٧٩٠	١٠٢٠	١٠٤٠	١٠٧٠	٩٧٠	١٠٠٠	١٠٩٠	١٠٣٠	١٠٧٠
١٠٦٠	٩٦٠	٨٣٠	١٠٧٠	١٠٢٠	١١٠٠	١٠٢٠	٧٦٠	٩٨٠	٨٨٠

أولاً: تحديد المدى الذى تتغير فيه الأطوال، أى تحديد أكبر وأقل قيمة فى البيانات وإيجاد الفرق بينهما. وبالنسبة للبيانات السابقة فإن أكبر طول هو ١٤٢٠ وأصغر طول هو ٥٧٠ وبهذا يكون المدى $٨٥٠ = ٥٧٠ - ١٤٢٠$

ثانياً: تحديد عدد الفئات. حيث أن عدد القيم أقل من ١٠٠٠ فإننا نستخدم معادلة Yule لإيجاد عدد الفئات كما يلى:

$$\text{عدد الفئات} = ٢,٥ = \sqrt[4]{١٠٠} \quad \text{فئة} = ٧,٩$$

ثالثاً: تحديد فترة الفئة، إذا استخدمنا فئات فإن

$$\text{فترة الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}} - \frac{٨٥٠}{٨} = ١٠٦,٢٥ \text{ متراً}$$

وإذا استخدمنا ٩ فئات فإن

$$\text{فترة الفئة} = \frac{\text{المدى}}{\text{عدد الفئات}} = \frac{٨٥٠}{٩} = ٩٤,٤٤ \text{ متراً}$$

ولتجنب وجود كسور عشرية فى مراكز الفئات فإنه من الأفضل أن تكون فترة الفئة رقماً صحيحاً وليكن ١٠٠ متراً. وكذلك فإنه من الملائم اختيار مراكز الفئات عند ٥٩٥، ٦٩٥، ٧٩٥، ولهذا فإن الفئات من الممكن أن تكون ٥٥٠ - ٦٤٠، ٦٤٠ - ٦٥٠، ٦٥٠ - ٧٤٠.. والجدول التالى يوضح عملية تكوين الجدول التكرارى.

جدول رقم (٢-١١)
جدول التوزيع التكرارى لأطول ١٠٠ رافد نهري

الفترة الفئة (الطول بالمتر)	حدود الفئة	مركز الفئة	الحزم	التكرار (ك)
٦٤٠-٥٥٠	٦٤٥-٥٤٥	٥٩٥	/	١
٧٤٠-٦٥٠	٧٤٥-٦٤٥	٦٩٥	//	٢
٨٤٠-٧٥٠	٨٤٥-٧٤٥	٧٩٥	///	٩
٩٤٠-٨٥٠	٩٤٥-٨٤٥	٨٩٥	////	٢٢
١٠٤٠-٩٥٠	١٠٤٥-٩٤٥	٩٩٥	/////	٣٣
١١٤٠-١٠٥٠	١١٤٥-١٠٤٥	١٠٩٥	//////	٢٢
١٢٤٠-١١٥٠	١٢٤٥-١١٤٥	١١٩٥	//////	٨
١٣٤٠-١٢٥٠	١٣٤٥-١٢٤٥	١٢٩٥	//////	٢
١٤٤٠-١٣٥٠	١٤٤٥-١٣٤٥	١٣٩٥	//////	١
	المجموع الكلى للمفردات = مجموع التكرارات			١٠٠

ويلاحظ أننا استخدمنا فى كل من الجدولين رقم (٢-٧) ورقم (٢-٨) فئات ذات فترات متساوية. ويطلق على التوزيع التكرارى من هذا النوع اسم «التوزيع المنتظم». ويفضل فى كل الحالات استخدام فئات متساوية وذلك تسهيلاً للعمليات الحسابية. وهناك نوع آخر من التوزيعات التكرارية يسمى بالتوزيع غير المنتظم، وفيه تكون أطوال الفئات غير متساوية وذلك بسبب تركيز عدد كبير من مفردات البيانات فى مدى ضيق نسبياً وانتشاراً العدد الأصغر منها على مدى واسع. ومن أمثلة هذا النوع من التوزيعات توزيع ملكية الأرض حسب فئات المساحة.

وتجدر الإشارة إلى أنه يوجد بصفة عامة نوعان من جداول التوزيع التكرارى حسب نوع التوزيع هما: ما يعرف «بالجدول المفتوح» وفيه يكون الحد الأدنى للفئة والحد الأعلى للفئة الأخيرة غير محدد. وهو فى هذه الحالة يكون مفتوح من طرفيه. ويكون الجدول مفتوحاً من طرفه الأعلى إذا كانت بداية الفئة الأولى غير

محددة. ويكون مفتوحاً من طرفه الأدنى إذا كانت نهاية الفئة الأخيرة غير محددة ويكون مفتوحاً من طرفه الأدنى إذا كانت نهاية الفئة غير محددة. أما النوع الثاني من الجداول فيعرف باسم «الجدول المقفول» وفيه يكون الحد الأدنى للفئة الأولى والحد الأعلى للفئة الأخيرة محددين كما في الجدولين رقم (٢-٨) و (٢-١١).

التوزيع التكرارى النسبى Relative Frequency Distribution

يعرف التكرار النسبى لفئة بأنه عبارة عن تكرار الفئة مقسوماً على التكرار الكلى لجميع الفئات، وعادة يعبر عنه كنسبة مئوية. ويستخدم التكرار النسبى فى حالة مقارنة توزيعين تكراريين يختلفان فى عدد مفردات كل منهما إذ أن مقارنة التكرارات المطلقة لكل فترة فى التوزيعين تكون غير صحيحة بسبب اختلاف حجم عينة كل منهما. فعلى سبيل المثال، فإن التكرار النسبى للفئة ٥٥٠ - ٦٤٠ فى الجدول رقم (٢-١١) هو $\frac{1}{3} = ٠.٣٣٣$ وإذا استبدلنا التكرارات فى الجدول التكرارى (جدول رقم ٢-١١) بما يقابلها من التكرارات النسبية فإن الجدول الناتج يسمى بالتوزيع التكرارى النسبى أو جدول التكرارات النسبية كما يبدو من الجدول التالى:

جدول رقم (٢-١٢)

التوزيع التكرارى النسبى لأطوال ١٠٠ رافد نهري

مركز الفئة	التكرار (ك)	التكرار النسبى
٥٩٥	١	$١ + ١٠٠ = ٠.٠١ (١)$
٦٩٥	٢	$٢ + ١٠٠ = ٠.٠٢ (٢)$
٧٩٥	٩	$٩ + ١٠٠ = ٠.٠٩ (٩)$
٨٩٥	٢٢	$٢٢ + ١٠٠ = ٠.٢٢ (٢٢)$
٩٩٥	٣٣	$٣٣ + ١٠٠ = ٠.٣٣ (٣٣)$
١٠٩٥	٢٢	$٢٢ + ١٠٠ = ٠.٢٢ (٢٢)$
١١٩٥	٨	$٨ + ١٠٠ = ٠.٠٨ (٨)$
١٢٩٥	٢	$٢ + ١٠٠ = ٠.٠٢ (٢)$
١٣٩٥	١	$١ + ١٠٠ = ٠.٠١ (١)$

التوزيع التكرارى المتجمع Cumulative Frequency Distribution

يسمى مجموع التكرارات المقابلة لجميع القيم الأقل من الحد الأعلى لفئة معينة بالتكرار المتجمع إلى هذه الفئة والمتضمن تكرارها أيضاً. وعلى سبيل المثال ففي الجدول رقم (٢-١١) فإن التكرار المتجمع فى الفئة ٨٥٠ - ٩٤٠ والمتضمن تكرارها أيضاً هو $1 + 2 + 9 + 22 = 34$. وهذا يعنى أن ٣٤ رافداً نهرياً تقل أطوالاً عن ٩٤٠ متراً. والجدول الذى يمثل التكرارات المتجمعة يسمى بجدول التوزيع المتجمع أو جدول التكرارات المتجمعة. كما هى الحال فى الجدول رقم (٢-١٣).

وفى بعض الأحيان قد يكون من المرغوب فيه الحصول على التوزيع التكرارى المتجمع لجميع القيم الأكبر أو المساوية للحد الأدنى لكل فئة. ويسمى التوزيع فى هذه الحالة بالتوزيع المتجمع على أساس «أكبر من» أو التكرار المتجمع.

جدول رقم (٢-١٣)

التوزيع التكرارى المتجمع الصاعد والهابط لأطوال ١٠٠ رافد نهري

التوزيع التكرارى المتجمع الهابط		التوزيع التكرارى المتجمع الصاعد		التوزيع التكرارى	
التكرار المتجمع	حدود الفئة	التكرار المتجمع	حدود الفئة	التكرار	الفئة
١٠٠	أكبر من ٥٥٠	١	أقل من ٦٤٠	١	٦٤٠-٥٥٠
$99 = 1 - 100$	أكبر من ٦٥٠	$2 = 1 + 1$	أقل من ٧٤٠	٢	٧٤٠-٦٥٠
$97 = 2 - 99$	أكبر من ٧٥٠	$12 = 1 + 9$	أقل من ٨٤٠	٩	٨٤٠-٧٥٠
$88 = 9 - 97$	أكبر من ٨٥٠	$34 = 12 + 22$	أقل من ٩٤٠	٢٢	٩٤٠-٨٥٠
$66 = 22 - 88$	أكبر من ٩٥٠	$67 = 34 + 33$	أقل من ١٠٤٠	٣٢	١٠٤٠-٩٥٠
$33 = 33 - 66$	أكبر من ١٠٥٠	$89 = 22 + 67$	أقل من ١١٤٠	٢٢	١١٤٠-١٠٥٠
$11 = 22 - 33$	أكبر من ١١٥٠	$97 = 8 + 89$	أقل من ١٢٤٠	٨	١٢٤٠-١١٥٠
$3 = 8 - 11$	أكبر من ١٢٥٠	$99 = 2 + 97$	أقل من ١٣٤٠	٢	١٣٤٠-١٢٥٠
$1 = 2 - 3$	أكبر من ١٣٥٠	$100 = 1 + 99$	أقل من ١٤٤٠	١	١٤٤٠-١٣٥٠

الهابط (النازل). بينما يسمى التوزيع المتجمع على أساس «أقل من» بال تكرار المتجمع الصاعد وهو يمثل مجموع القيم التي تقل في قيمتها عن الحد الأعلى للفتة، أى تساوى تكرار الفتات السابقة لفتة ما مضافاً إليها تكرار الفتة نفسها.

وكما ذكرنا سابقاً يمكن استبدال التكرار المتجمع المطلق في الجدول السابق إلى تكرار متجمع نسبي (أو تكرار متجمع مئوى) Relative Cumulative Frequency وذلك بقسمة التكرار المتجمع في كل فتة على التكرار الكلى. فمثلاً.

$$\text{التكرار المتجمع النسبي للأطوال أقل من } 1040 \text{ متراً. هو } \frac{67}{100} = 67\%$$

وهذا يعنى أن 67% من الروافد أطوالها أقل من 1040 متراً. وإذا وضعت التكرارات المتجمعة النسبية في جدول فإنها تسمى بالتوزيع التكرارى المتجمع النسبى أو بالتوزيع المتجمع للنسب المئوية.

الفصل الثالث

العرض البياني للبيانات الإحصائية

يعتمد أسلوب العرض البياني على ترجمة المعلومات وتلخيص البيانات الإحصائية (المبوبة وغير المبوبة) ووضعها في صورة أشكال بيانية أو في هيئة رسوم تصويرية تسهل فهم واستيعاب الخصائص والاتجاهات والعلاقات المختلفة والمتشابهة للظواهر الجغرافية موضع الدراسة. وتبعاً لذلك فإن الأشكال والرسوم البيانية تعد خير وسيلة للتعبير وتوصيل المعلومات، كما يمكن أن نعتبرها لغة ثانية يشرح بها الباحث موضوع بحثه دون أن يجهد القارئ أو المشاهد في استخلاص الحقائق من الجداول والأرقام. وتمتاز الأشكال والرسوم البيانية والتصويرية بأنها تعطي فكرة سريعة للناظر إليها من أول وهلة، بينما لا يظهر هذا الأثر إذا مانظرنا إلى بيانات رقمية في جدول أو إحصائية. لكل ذلك فإننا - بحق - يمكن أن نقول أن العرض البياني هو روح البيانات وسبيل إلى الوصول إلى ما تخبؤه من معلومات.

العرض البياني للبيانات الخام (غير المبوبة)

تختلف وتتعدد طرق وأساليب العرض البياني للبيانات الخام (غير المبوبة)، ولكنها تنحصر في ثلاث مجموعات رئيسية هي: مجموعة الطرق البيانية لتمثيل التغير في مكونات عناصر الظاهرة والمجموع الكلي لها والطرق البيانية بالرسوم التصويرية. وفيما يلي دراسة تفصيلية لكل مجموعة على حدة.

أولاً: الطرق البيانية لتمثيل العلاقات بين كميات المتغيرات:

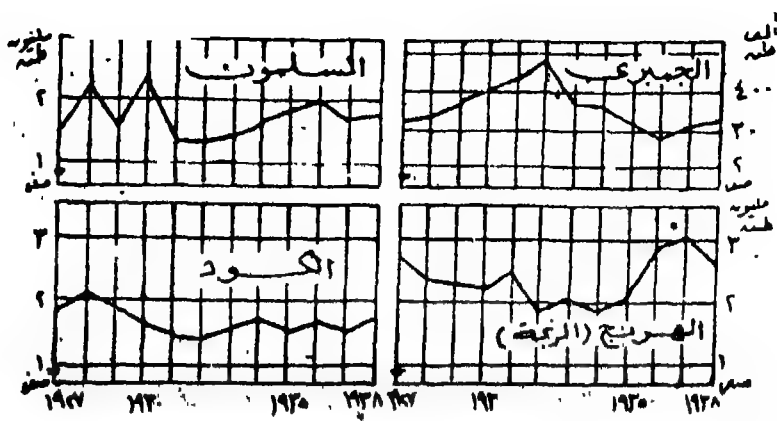
يمكن تقسيم هذه الطرق، حسب البيانات الإحصائية المتاحة للمتغيرات موضع

الدراسة إلى قسمين: القسم الأول يختص ببيان العلاقات بين الكميات المختلفة لعدة متغيرات، والقسم الآخر يهتم بإظهار تطور مقدار (قيم) المجموع الكلى لمتغير واحد أو عدة متغيرات. ويتضمن القسمان مجموعة غير قليلة من أساليب التمثيل البياني سنتعرض لشرح أكثرها أهمية في تحقيق الغرضين السابقين، مثل الخطوط البيانية الحسائية واللوغاريتمية، أشكال الإنتشار، الأعمدة البيانية البسيطة، والرسوم البيانية الحجمية والرسوم الدائرية، فيمايلي:

١ - اخطط البياني البسيط Simple-Line graph

تستخدم الخطوط البيانية البسيطة في تمثيل التغير من فترة إلى أخرى للظاهرة الواحدة أو لبيان علاقة متغيرين وغالبا ما يكون أحد هذين المتغيرين هو الزمن الذي يعتبر متغيراً مستقلاً. ويبين التغير أو العلاقة بمنحنى يظهر منه ضعف أو شدة التغير من فترة إلى أخرى ويكون ذلك ما يعرف بالسلسلة الزمنية، أو يوضح اتجاه العلاقة بين متغيرين أو أكثر.

ولقد جرت العادة عند التمثيل البياني للسلاسل الزمنية التي توضح مقادير أو كميات المتغيرات لعدد من السنوات أن يكون المحور الأفقى «الزمنى» ممثلاً للمتغير المستقل «الزمنى» والمحور الرأسى «الصادى» للمتغير التابع الظاهرة موضع الدراسة. وفي ضوء البيانات المتاحة يختار مقياس رسم ملائم لأبعاد المسطح المخصص لعملية التمثيل البياني حتى يمكن توقيع كل قيم المتغير التابع على الرسم. فيقسم المحور الرأسى إلى وحدات حسائية بادئين بالصفير ومنتهين بقيمة أكبر من أكبر قيمة تمثل المتغير التابع. ويختار كذلك مقياس مناسب للمتغير المستقل «الزمن» على المحور الأفقى ويجب مراعاة عدم وجود تفاوت كبير فى الأبعاد القياسية للمتغيرين حتى لا يؤدي ذلك إلى نقص الدقة فى تمثيل بيانات المتغيرين. ويتم رسم المنحنى من خلال توقيع جميع القيم على الرسم فى شكل نقط تحدد كل منها باحداثين (أى على حسب كل نقطة عن المحورين) ثم توصل مواقع القيم فتعطى لنا الشكل المطلوب (الخط البياني) كما فى الشكل رقم (٣-١).



شكل رقم (٣-١)

تطور إنتاج الثروة السمكية في كندا في الفترة من ٢٧ - ١٩٣٨

(طريقة الخطوط البيانية البسيطة)

ويجب ملاحظة أنه عند رسم الخطوط البيانية لظاهرة متغيرة بانتظام أو تدريجياً أن يكون الخط البياني منحنياً، مثل الخط البياني الذي يوضح المتوسط الشهري لدرجة الحرارة خلال شهور السنة في مدينة الاسكندرية. أما إذا كانت الظاهرة لا تتغير بانتظام مثل كمية الأمطار أو كمية إنتاج أحد المحاصيل الزراعية فإنه يجب أن يكون التوصيل بين النقط على شكل خط منكسر. وفي بعض الحالات قد لا يبدأ المقياس على المحور الرأسى بالصففر ولكن يبدأ بقيمة أكبر تبعاً لأن البيانات المراد تمثيلها بيانياً بخط بياني تبدأ بقيمة بعيدة عن الصففر ولكن تقترب من بعضها بمرور الزمن، فإذا ما أخذنا وحدات قياسية تناسب أصغر وأكبر رقم بادئين بالصففر فإن ذلك سيؤدي إلى وجود فراغ كبير غير مستخدم يقع بين الصففر وأصغر رقم موجود مما يترتب عليه أن يجعل الخط البياني محصوراً في أعلى جزء من الرسم وهذا شيء غير مستحب أو مرغوب فيه. وللتغلب على ذلك فإننا نحاول أن نضغط المسافة على المحور الرأسى بين الصففر وأصغر قيمة بأن نكسر المحور الرأسى بخطين مائلين بعد نقطة الصففر على المحور الرأسى.

وفى بعض الأحيان إذا كان المطلوب تمثيل بيانات متغير واحد فى أكثر من مكان لابرار العلاقات والاختلافات المكانية لهذا المتغير، أو تمثيل بيانات متغيرين أو أكثر فى مكان واحد لابرار خصائص المتغيرات موضع البحث لهذا المكان، فإنه يمكن رسم وتوقيع هذه البيانات بأكثر من خط بيانى بسيط فى شكل بيانى واحد يعرف باسم « الخطوط البيانية المتعددة أو البوليجراف Polygraphs ». والقاعدة الأساسية لإنشاء البوليجراف تلخص فى رسم محورين رأسيين - محور أيسر وآخر أيسر - كل منهما يقسم إلى تقسيمات تختص ببيانات أحد المتغيرين.

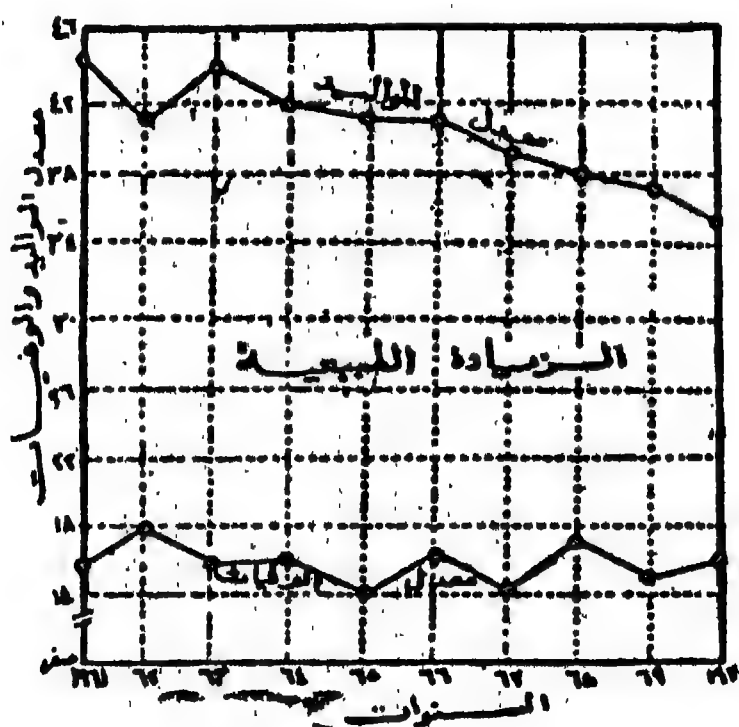
ويستخدم هذا النوع من الرسوم البيانية على نطاق واسع لتوضيح العلاقات بين بيانات العناصر المناخية، مثل بيان العلاقة بين درجة الحرارة وكمية التبخر أو بين كمية الأمطار وعدد الأيام الماطرة، أو بين بيانات أحد العناصر المناخية وبيانات ظاهرات جغرافية طبيعية أخرى مثل التضاريس والنبات، أو ظاهرات بشرية مثل عدد المصطافين أو عدد السياح فى دولة ما، أو كمية استهلاك المياه فى مدينة ما، أو حتى نوع العمل الذى يقوم به الأفراد فى مجتمع ما. وبصفة عامة فإن هناك علاقات عديدة لا حصر لها يمكن أن نوضحها باستخدام مثل هذا النوع من الرسوم البيانية.

٢ - الخطوط البيانية اللوغاريتمية Logarithmic graphs

فى حالة اذا كنا بصدد تمثيل سلسلة زمنية لبيانات تتفاوت القيم فيها تفاوتاً كبيراً أو تمثيل بيانات فى شكل معدلات النمو أو التغير السكانى من سنة لأخرى، أو معدل التغير فى الاستهلاك، أو معدلات التغير فى الدخل القومى، أو نسب تطور الدعم الحكومى للسلع والخدمات، أو نسب النقص .. والزيادة فى أى ظاهرة، فإنه يجب أن يقسم المحور الرأسى إلى وحدات لوغاريتمية بدلا من الوحدات الحساسة.

وتقوم فكرة التقسيم اللوغاريتمى على أخذ لوغاريتم الأعداد من ١ إلى ١٠ وجعلها أساساً لوحدة التقسيم اللوغاريتمى والتي تضرب فى كل مرة فى طول الدورة اللوغاريتمية المأخوذة طبقاً لطول المسافة الرأسية والأفقية المراد تمثيل الظاهرة عليها، وهى فى هذه الحالة تمثل دورة لوغاريتمية واحدة. وبعد ذلك يمكن أيضاً

وكما فى رسم الخطوط البيانية الحسائية وعلى حسب البيانات المتاحة يمكن أن يقسم المحور الرأسى فقط تقسيماً لوغاريتمياً لتوقع على أساسه معدلات التغير فى ظل الفترة الزمنية التى يقسم على أساسها المحور الأفقى تقسيماً حسائياً، يسمى ذلك بالتقسيم نصف لوغاريتمى. كما قد يقسم كل من المحورين الأفقى والرأسى تقسيماً لوغاريتمياً يطلق عليه اسم التقسيم اللوغاريتمى المزدوج، ويناسب ذلك البيانات التى تتكون من معدلات تغير أو نسب مئوية لمتغيرين مستقلين. وتبعاً لأهمية استخدام التقسيم اللوغاريتمى فى التمثيل البيانى فإنه يوجد حالياً ورق رسم بيانى خاص مقسم تقسيماً لوغاريتمياً إما على المحور الأفقى أو الرأسى أو على المحورين معاً (تقسيم مزدوج) شكل رقم: (٢-٣).

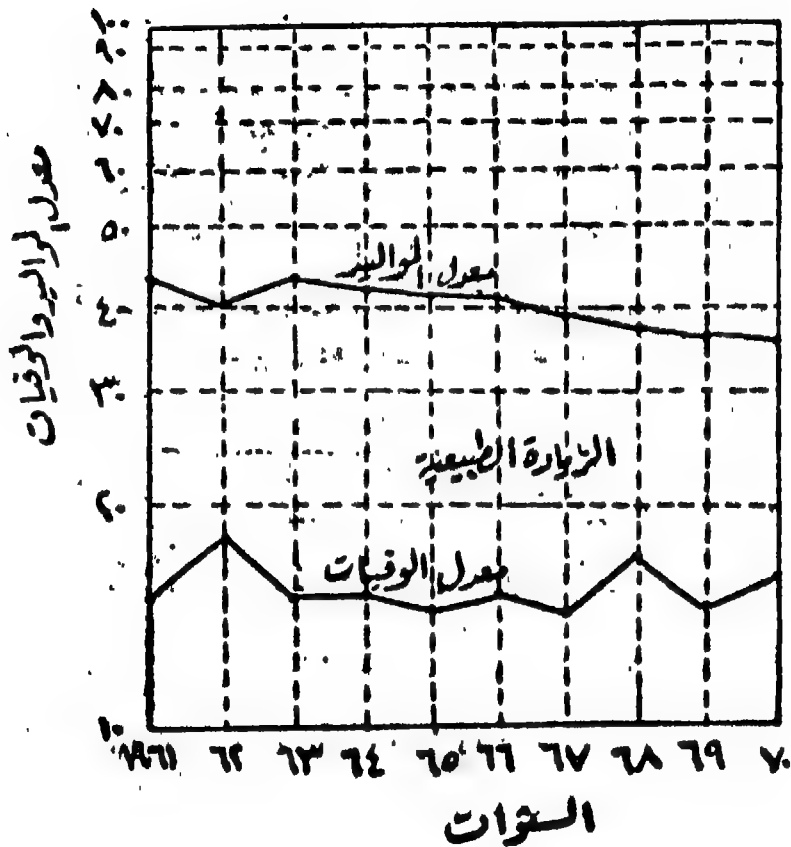


شكل رقم (٣-٣)

معدلات المواليد والوفيات ١٩٦١ - ١٩٧٠.

(خطوط بيانية بسيطة - لاحظ كسر المحور الرأسى بعد الصفر)

والشكلان رقم (٣-٣)، (٤-٣) يوضحان مقارنة بين الخطوط البيانية الحسائية واللوغاريتمية لمعدلات المواليد والوفيات في جمهورية مصر في الفترة من ١٩٦١ - ١٩٧٠، ومنهجات يتضح أن الخطوط البيانية اللوغاريتمية التي رسمت على رسم بياني نصف لوغريتمى لانتظر حدة التغير في كل من معدلات المواليد والوفيات التي أظهرتها الخطوط البيانية الحسائية وانعكس ذلك أيضاً على الزيادة الطبيعية للسكان المحسوبة من الرسم في كل من الشكلين.



شكل رقم (٣-٤)

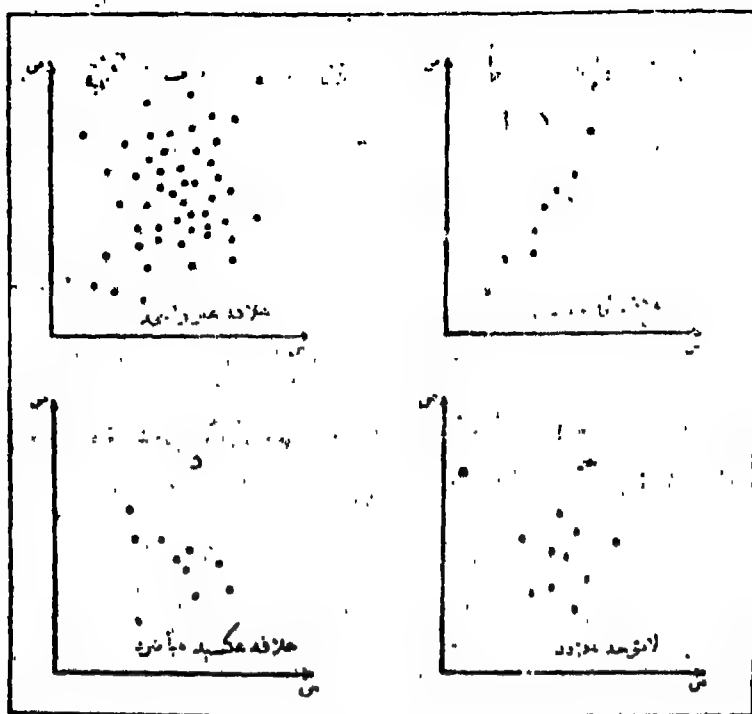
معدلات المواليد والوفيات في الفترة من ١٩٦١ - ١٩٧٠

(رسم بياني لوغاريتمى)

٣- أشكال الانتشار Scatter graphs

تبين أشكال الانتشار - بطريقة تقريبية - العلاقة أو مدى الترابط واتجاهاته بين متغيرين أو أكثر بفعل خصائص ومعالم معينة تربط بينهما. وقد توحى هذه الأشكال بعدم وجود ارتباط بين خصائص أحد المتغيرات ومعالم المتغير أو المتغيرات الأخرى قيد البحث.

وتتمثل اتجاهات الترابط أو العلاقة بين المتغيرات فى أربعة أشكال (شكل رقم ٣-٥): الشكل الأول منها «أ» يوضح علاقة طردية (موجبة) مباشرة بين المتغيرين، بمعنى أن تزايد قيمة أحد المتغيرين يصاحبها تزايد فى قيمة المتغير الآخر. مثال ذلك علاقة أو ارتباط أجر العمال فى أحد المصانع بكمية الوحدات المنتجة فى هذا المصنع. فإذا ارتفع أجر هؤلاء العمال ازدادت كمية الوحدات المنتجة. وبالمثل ارتباط انتاج المحاصيل الشتوية على الساحل الشمالى لمصر بكمية الأمطار، بمعنى أنه كلما ازدادت كمية الأمطار الساقطة كلما ازدهرت هذه المحاصيل وزاد انتاجها والعكس صحيح. ويتمثل الشكل الثانى (د) للعلاقة فى الارتباط العكسى أو السلبى المباشر بين المتغيرين، بمعنى أن الاختلافات أو التطور فى أحد المتغيرين يصاحبه اختلاف وتطور معاكس للتغير الآخر. مثال ذلك العلاقة الشهيرة بين كثافة السكان والبعد عن قلب المدينة. فمن المعروف أن كثافة السكان تزداد بالقرب (أو بتناقض المسافة) من قلب المدينة والعكس صحيح. وبينما يتمثل الشكل الثالث (ب). للعلاقة بين المتغيرات فى وجود علاقة غير واضحة بين أحد المتغيرات والمتغير الآخر بمعنى أن الترابط بينهما لا يوجد له اتجاه منتظم أو محدد، فإن الشكل الرابع (ج) للترابط يتمثل فى عدم وجود علاقة ارتباطيه على الإطلاق بين المتغيرات قيد البحث.



شكل رقم (٣-٥)

أشكال الانتشار لتوضيح اتجاهات الترابط بين المتغيرات

ويمكن الحصول على شكل الانتشار الذي يحدد نوع ودرجة العلاقة بين المتغيرات عن طريق توقيع بيانات المتغيرات على رسم بياني واحد يتألف من إحداثيين أحدهما أفقي والآخر رأسي؛ يخصص أحدهما لتوقيع بيانات المتغير الثاني. فإذا ما وقعت القيم الممثلة للمتغيرين في مجال انتشار متقارب أو على امتداد خط مستقيم دل ذلك على وجود علاقة قوية (طردية أو عكسية). أما إذا أظهر شكل الانتشار تباعداً طفيفاً للقيم ولكن حول خط مستقيم دل ذلك على وجود علاقة ضعيفة. وإذا ما سجل شكل الانتشار تباعداً كبيراً للنقط بحيث يتعذر معه أن تقع على خط مستقيم فإن ذلك يدل على عدم وجود علاقة ترابط بين المتغيرين.

٤ - الأعمدة البيانية البسيطة Bar graphs

تعد طريقة الأعمدة البيانية من أيسر طرق التمثيل البياني التي تستخدم للمقارنة بين الكميات لظاهرة واحدة أو عدة ظواهر، وعادة ماتسمى رسومها البيانية باسم Columnar diagrams. وتتألف هذه الرسوم من أعمدة ذات عرض متساوى وطول يتناسب مع الكميات التي تمثلها حسب مقياس الرسم المختار. ويمكن رسم هذه الأعمدة أما رأسيًا أو أفقيًا في أشكال بيانية قائمة بذاتها، وتعتبر الأعمدة الأفقية أفضل من حيث سهولة قراءتها، أما الأعمدة الرأسية فلها ميزة أخرى وهي سهولة المقارنة.

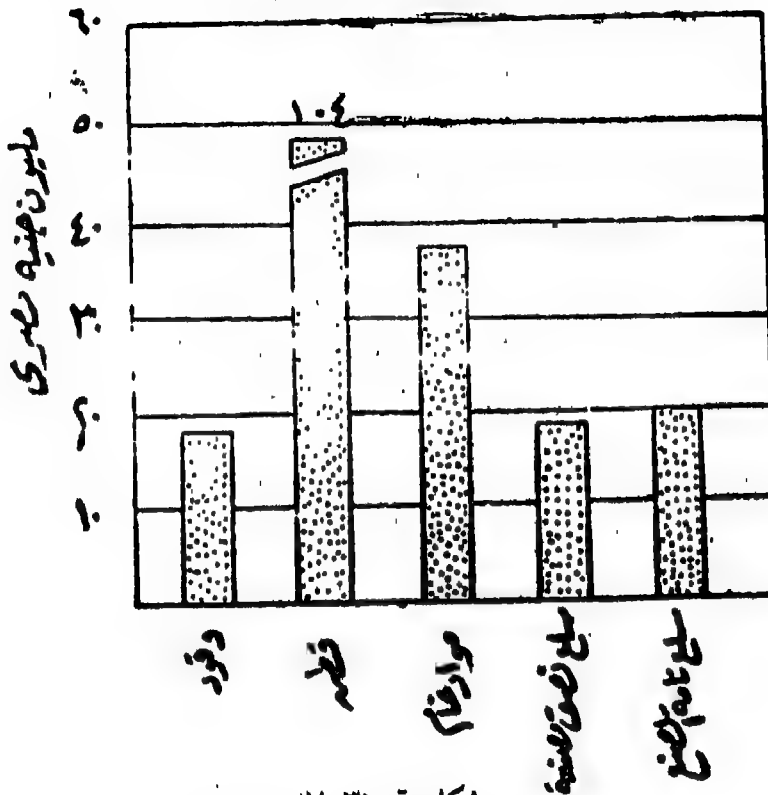
وتعتمد طريقة الأعمدة البيانية البسيطة في تمثيل البيانات الوصفية على اظهار كميات الزيادة والنقص في بعض الظواهر مثل معدلات المواليد والوفيات أو تمثيل التطور في أي سلسلة زمنية خاصة بالإنتاج أو الاستثمار أو حجم المشتريات أو الدخل أو عدد السكان خلال فترات زمنية معينة. وفي هذه الحالة نرسم محورين أحدهما محور رأسي يقسم إلى أقسام متساوية تبين الكميات والآخر محور أفقي يقسم أيضاً إلى أقسام متساوية حسب الفترات الزمنية أو الصفات المميزة للظاهرة كالحالات التعليمية أو الإجتماعية أو فئات السن .. الخ وما هو جدير بالذكر أنه عند أخذ المسافات المثلثة لقواعد الأعمدة على المحور الأفقي يجب أن تكون متساوية وعلى أبعاد متساوية أيضاً، وذلك بطريقة تلائم المساحة من لوحة الرسم المخصصة للتمثيل البياني وعدد الأعمدة المراد رسمها (شكل رقم ٣-٦). وفي حالة الفترات الزمنية غير المنتظمة فإن المسافات بين كل عمود وآخر يجب أن يتناسب مع الأبعاد الزمنية للفترة المراد تمثيلها بيانياً. كما يجب في كل الحالات أن يبدأ المقياس على المحور الرأسي من الصفر وينتهي برقم أعلى من أكبر قيمة من قيم الظاهرة موضع البحث. إلا أنه في كثير من الحالات نجد بين قيم الظاهرة المراد تمثيلها بطريقة الأعمدة البسيطة قيمة أو قيمتين متطرفين أو شاذتين تفوق بقية قيم الظاهرة بما يؤدي إلى وجود تفاوت كبير لهذه القيم. وبالتالي يؤدي ذلك إلى



شكل رقم (٣-٦)
قيمة منتجات مصايد الأسماك الكندية
في الفترة ١٩١٧ - ١٩٣٨ (بالمليون دولار)

اختلاف كبير في طول الأعمدة، بل أنه في بعض الأحيان يصبح من الصعب تمثيل القيم بأخذ مقياس رسم على المحور الرأسي يلائم هذه القيم المتفاوتة. فمثلاً إذا كانت لدينا كمية أكبر مائة مرة من كمية أخرى، فإنها تتطلب رسم عمود أطول مائة مرة من عمود الكمية الأصغر. وهذا يضطرننا طولها الكبير الصغيرة بأعمدة صغيرة جداً، وإما أن نرسم أعمدة قد يضطرننا طولها الكبير جداً إلى تقطيعها إلى قطاعات توضع بجوار بعضها البعض. ولو أن كل هذا الضخام لا ينقل الصورة الصحيحة لتمثيل هذه الكميات، ومما لذلك من تقليل من أهمية هذا الأسلوب في التمثيل البياني. وللتغلب على هذه المشكلة يستحسن قطع المحور الرأسي الموجود عليه لقياس الكميات وجعل الجزء الأسفل منه يبدأ من الصفر حتى قيمة أعلى من الكمية الصغيرة، أما الجزء الأعلى فيبدأ من رقم أقل من الكمية الكبيرة وينتهي برقم أعلى منها مع ثبات طول المقياس في الجزئين. وفي بعض الأحيان تكسر الأعمدة التي تمثل قيما متطرفة ويكون ذلك بالتخلص من

الارتفاعات العادية للقيم الأخرى وقد يكون كسر الأعمدة رأسياً عن طريق وضع خطين متوازيين مائلين عند نهاية الارتفاع المراد تحديده والذي يناسب الشكل وليدل على أن العمود بقية، ولكن مساحة ورقة الرسم لا تسمح باظهارها، ويجب أن نكتب أعلى هذا العمود بالذات الكمية الحقيقية التي يمثلها (شكل رقم ٣-٧). وعلى الرغم من ذلك فإن هذه الطريقة لا يمكن الاستفادة بها في حالة المقارنة لأنها لا تظهر الفرق بين الكميات كحقيقتها.



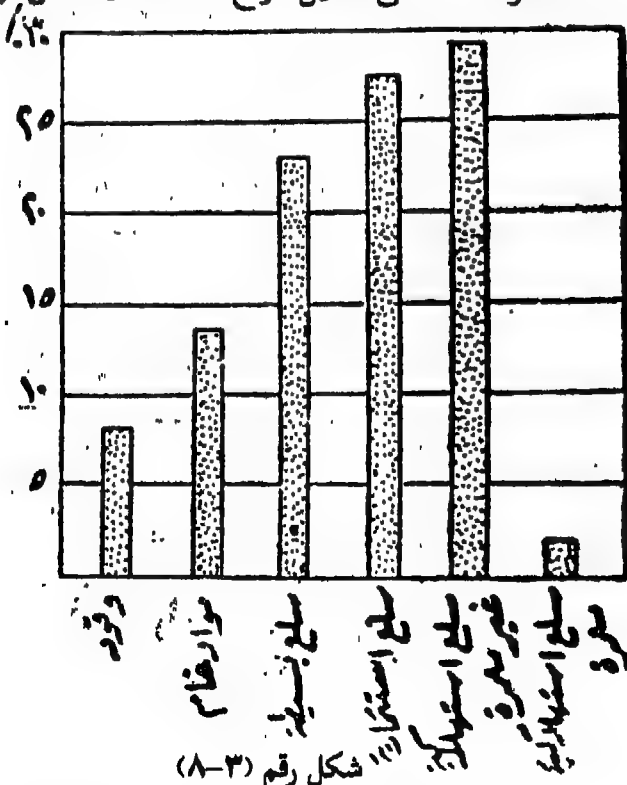
شكل رقم (٣-٧)

قيمة الصادرات المصرية في عام ١٩٦٣، ٦٢

بالأعمدة البيانية البسيطة (لاحظ كسر العمود الثاني)

وهناك نوع آخر من الأعمدة البيانية يصلح في اظهار الأهمية النسبية لكمية الظاهرة يسمى بالأعمدة النسبية Proportional Bargraph والتي ترسم لتمثيل أعداد السكان أو الإنتاج المعدني أو حركة الصادرات والواردات في الموانئ. وتتميز

طريقة الأعمدة النسبية بسهولة رسمها من ناحية التصميم، وكذلك بسهولة القراءة من الناحية المرئية. وتتلخص طريقة رسم الأعمدة النسبية في أن نبدأ أولاً باستخراج النسبة المئوية للكميات التي نريد تمثيلها بالنسبة للمجموع الكلى للكميات، مثل نسبة وزن المجموعات الرئيسية للواردات المصرية في سنة ١٩٦٢، ويخصص المحور الأفقى لتعيين المجموعات المختلفة للواردات، أما المحور الرأسى فيخصص للأوزان المناظرة لكل مجموعة ويقسم إلى أقسام متساوية تبين النسبة المئوية للأوزان مبتدئين بالصفير ومنتهين بنسبة أعلى من أعلى النسب المراد تمثيلها كما في الشكل رقم (٨-٣). ويحسن عند رسم هذا النوع من الأعمدة أن نختار له نوع من التظليل المصمت (كاللون الأسود المصمت) أو نستخدم نمط التظليل النقطى وذلك لأن استخدام نمط الخطوط المائلة في تظليل فراغ الأعمدة يتضمن نوعاً من خداع البصر.



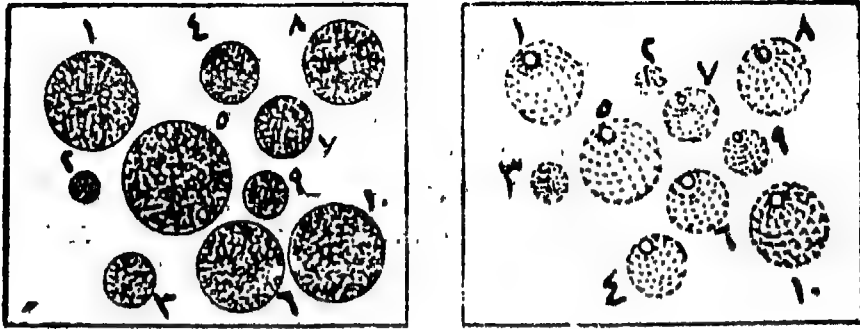
أوزان المجموعات الرئيسية للواردات المصرية ١٩٦٢ / ١٩٦٣
الأعمدة البيانية النسبية

٥- الرسوم البيانية الحجمية Three- dimensional graphs

إذا كانت البيانات المراد تمثيلها بيانياً ذات مدى عظيم جداً في القيم أو الكميات، فالتدخل البعد الثالث الذي يترتب عليه استخدام رسوم بيانية حجمية تتناسب أحجامها مع مقدار الكميات التي تمثلها. ومن أهم هذه الرسوم البيانية الكرات Spheres والمكعبات Cubes. وعلى الرغم من مميزات هذا النوع من الرسوم البيانية فإن هناك بعض من المثالب التي يمكن إجمالها في: أن رسم الرسوم الحجمية ليس أمراً سهلاً بل يتطلب جهداً وعملاً إضافياً حتى يبدو الشكل الحجمي واضحاً، أو بمعنى آخر أن تعطى الكرة أو المكعب الشكل الحجمي بأبعاده الثلاثة على سطح لوحة الرسم المستوي. وعلى الرغم من أن العلاقة بين أحجام الأشكال والكميات التي تمثلها صحيحة رياضياً إلا أنه ليس من السهل تقدير أحجام هذه الأشكال بمجرد النظر إليها عكس الأعمدة البيانية.

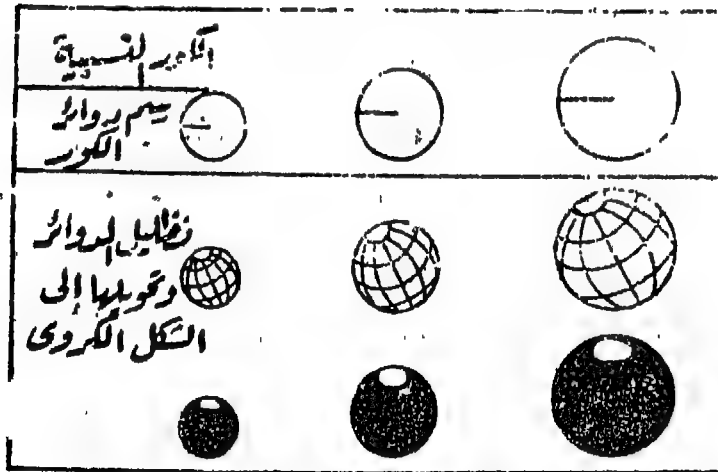
وفي حالة استخدام الرسوم البيانية الحجمية لتمثيل ومقارنة كميات عظيمة التفاوت والاختلاف فإن حجم هذه الأشكال تتناسب مع مكعب نصف القطر (في حالة الكرات) أو مع مكعب طول الضلع (في حالة المكعبات) فالكرة الأكبر عشرة مرات من كرة أخرى سوف تمثل كمية أكبر ألف مرة (١٠^٣) من الكمية التي تمثلها الكرة الأخرى. وكما هو متبع في طريقة رسم الدوائر البيانية، فإننا نستخرج أولاً الجذور التكعيبية للكميات، ونعتبر هذه الجذور التكعيبية أنصاف أقطار للدوائر التي سنعطيهها شكل الكرات، أو نعتبرها طول المكعبات المراد رسمها. وفي حالة رسم الكرة نبدأ أولاً برسم دائرة عادية ثم نعطيها الشكل الحجمي، أما أن نجعلها تمثل شكل «الكرة الأرضية» وذلك برسم شبكة رمزية من دوائر العرض وخطوط الطول فوق الدائرة المفرغة والتي ستبدو في النهاية على شكل كرة مجسمة، وأما أن نطمس كل مساحة الدائرة باللون الأسود مع ترك مساحة بيضاء في أعلى الكرة بحيث تبدو كالنور الساطع في أعلى الكرة كما في الشكل رقم (٣-٩ب)، أما المكعبات فهناك نوعان منها: نوع يبدو على شكل الدوالاب وفيه يكون طول الجوانب مساوية لنصف طول الوجهة، والنوع الآخر يبدو متساوي الأضلاع والارتفاعات وتكون أطوال الجوانب والوجهة متساوية. وبعد أحسن شكل للمكعب

هو الذى يكون فيه طول ضلع جوانبه $\frac{2}{5}$ طول ضلع واجهته، بحيث تميل هذه الجوانب من ٣٠ إلى ٥٠ من الخط الأفقى وتكون جوانب المكعب على يمين الناظر للرسم البياني.



شكل رقم (٣-١٩)

الرسم البياني الحجمية «كرات» والمساحية «دوائر»
لاحظ اختلاف حجم الكرات عن مساحة الدوائر لنفس البيانات



شكل رقم (٣-٩ب)

الكرات النسبية بيانات سكان المدن المصرية لعام ١٩٦٦

وكمثال لتطبيق طريقة الرسوم الحجمية يمكن تمثيل عدد سكان كل من القاهرة والاسكندرية والجيزة (أكبر المدن المصرية) بالكرات أو المكعبات كما هي الحال في الشكل رقم (٣-٩ ب) الذى يعتمد على البيانات للجدول التالى:

جدول رقم (٣-١)

عدد سكان أكبر المدن المصرية عام ١٩٦٦

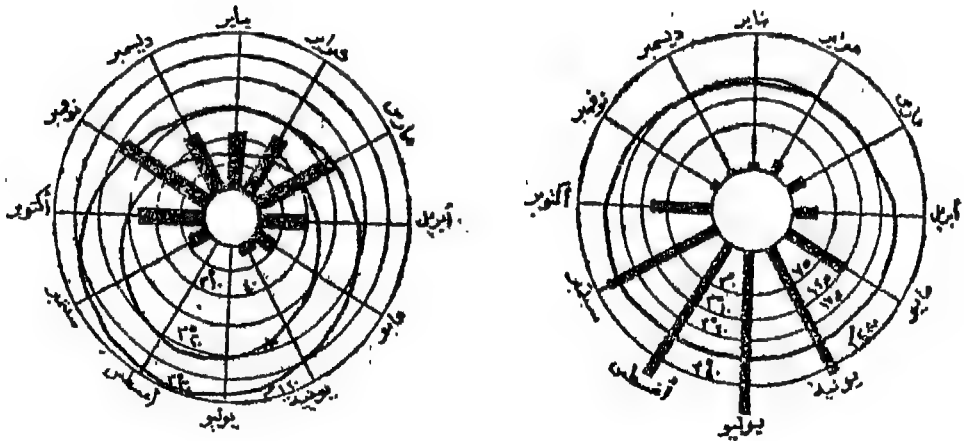
عدد السكان المدينة	عدد السكان (بالآلاف)	الجزر التكميى	طول قطر الكرة أو ضلع المكعب (الجدول التكميى ٢٠٠)
القاهرة	٤,٢٢٠	١٦١,٦٠	٠,٨ سنتيمتر
الاسكندرية	١,٨٠١	١٢١,٦٤	٠,٦ سنتيمتر
الجيزة	٥٧٠	٨٢,٩٧	٠,٤ سنتيمتر

٦- الرسوم البيانية الدائرية Circular (Colck) graphs

تستخدم طريقة الرسوم البيانية الدائرية (Polar Chart) لتمثيل بيانات المتغيرات المستمرة، مثل بيانات درجة الحرارة والمطر لمختلف محطات الأرصاد الجوية فى بلدما، أو توزيع متوسط نسبة العاملين فى أقسام أحد المحلات التجارية الكبرى على شهور السنة، أو توزيع العمالة فى مختلف النشاط الزراعى على مدار شهور السنة. وتتفوق هذه الطريقة على طريقة الخطوط البيانية البسيطة، فى أن نهايات الرسم البيانى (الطرف الأيمن والطرف الأيسر) فى الطريقتين الأخيرتين يقطع استمرار تطور الظاهرة أو يخفى اتجاهها العام على مدى فترة التوزيع.

وتعتمد طريقة الرسوم الدائرية على رسم دائرة وتقسيمها إلى ١٢ قسماً تبعاً لعدد شهور السنة بحيث يكون نصيب كل شهر ٣٠ درجة من درجات الدائرة (٣٦٠ درجة)، ثم يرسم من مركز هذه الدائرة كميات المتغير قيد البحث، مثل درجات الحرارة المختلفة، على أن نراعى أن يشمل ذلك أقل الكميات (الدرجات) وأعلاها. فإذا كان مقدار أو كمية المتغير لاتقل عن الصفر ففى هذه الحالة ج

يصبح مركز الدائرة هو الصفر، أما إذا كان هذا المقدار أقل من الصفر، فيصبح مركز الدائرة في هذه الحالة ممثلاً لرقم يقل عن الصفر في حين يمثل رقم الصفر نفسه بدائرة منفصلة. فمثلاً إذا كانت لدينا بيانات عن المتوسطات الشهرية لدرجة الحرارة وكمية المطر لمحطات جوية تتميز بفصلية توزيعه هذين العنصرين المناخيين فإنه يمكن تمثيل هذه البيانات برسم بياني دائري عن طريق رسم أنصاف أقطار على بعد من مركز الدائرة الرئيسية «الخارجية» يمثل كل منها بيانات كل عنصر تبعاً لقيمتها. ثم نصل بين نهايات هذه الأنصاف أقطار بخط منحنى وبذلك تتكون لدينا شبه دائرة محيطها متعرج يمثل عنصر درجة الحرارة كما نرسم على أنصاف الأقطار أعمدة بيانية بسيطة تمثل أطوالها كميات المطر الساقطة في كل شهر من شهور السنة (شكل رقم ٣-١٠).



شكل رقم (٣-١٠)

الرسوم بيانية الدائرية لتمثيل عنصرى
درجة الحرارة والمطر في كلكتا «الهند» وملقا «إسبانيا»

ويمكن تقسيم شبه الدائرة الناتجة عن تمثيل درجة الحرارة إلى شرائح تمثل كل شريحة منها فصلاً حرارياً مميزاً مثل الفصل الحار والفصل البارد أو الفصل الذى تسمح فيه درجة الحرارة بنمو النباتات مثلاً. كما يجوز عن طريق هذا الرسم تحديد اليوم الذى يبدأ فيه كل فصل من هذه الفصول الحرارية.

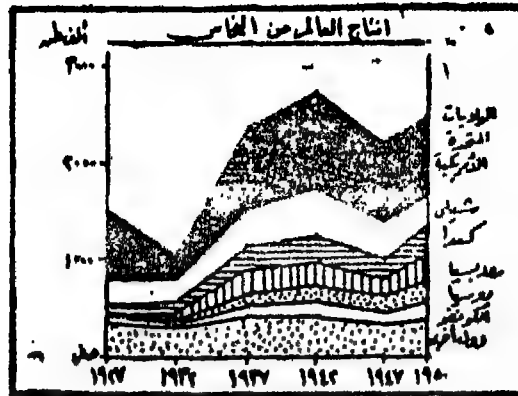
وعلى الرغم من تغلب طريقة الرسوم البيانية الدائرية على مشكلة انفصام توزيع البيانات المستمرة على الرسوم البيانية العادية «طريقة الخط البياني البسيط مثلاً»، إلا أن استخدامهما ليس شائعاً أو متداولاً مثل بقية الطرق البيانية بسبب أن تذبذب «ارتفاع وانخفاض» الخط البياني الذى يمثل بيانات المتغير يمكن ملاحظته مباشرة على الرسم البياني العادى، بينما يظهر ذلك فى الرسم الدائرى على شكل اقتراب أو ابتعاد عن مركز الدائرة، مما يتطلب دقة ومهارة خاصة عند قراءة وتفسير الشكل الناتج عنها.

ثانياً: الطرق البيانية لتمثيل التغير فى مكونات الظاهرة والمجموع الكلى لها:

إن الهدف الأساسى من هذه المجموعة من الطرق البيانية هو المقارنة بين قيم المكونات المختلفة لظاهرة ما بعضها البعض وبين كل منها والمجموع الكلى للظاهرة. وفيما يلى دراسة مختصرة توضح غرض وأسلوب إنشاء كل طريقة من الطرق الشائعة الاستخدام فى مجال العرض البياني للبيانات الاحصائية وهى: الرسوم البيانية المجمعة، الأعمدة البيانية المركبة، الأهرامات البيانية، الدوائر المقسمة الرسوم المثلثية.

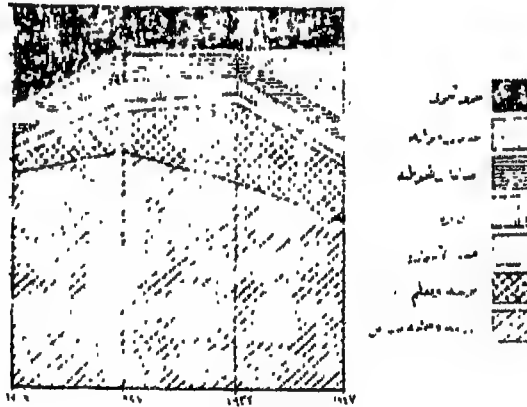
١ - الرسوم البيانية (المنحنيات) المجمعة Compound- line graphs

وهى عبارة عن خطوط بيانية تمثل التغير فى مجموع الظاهرة الواحدة على مدى فترة زمنية أو التغير فى مجموع الظاهرة وظاهرة (أو ظاهرات) أخرى، بحيث يمثل التغير فى أجزاء أو مكونات الظاهرة بخطوط بيانية بسيطة يمثل كل خط منها بداية القياس الخط اللاحق له، ثم تظلل المساحات المحصورة بين هذه الخطوط البيانية (شكل رقم ٣-١١)



شكل رقم (٣-١١)
إنتاج النحاس في العالم «بالآلاف الأطنان المترية»
«طريقة الرسوم البيانية المجمعة»

٪١٠٠

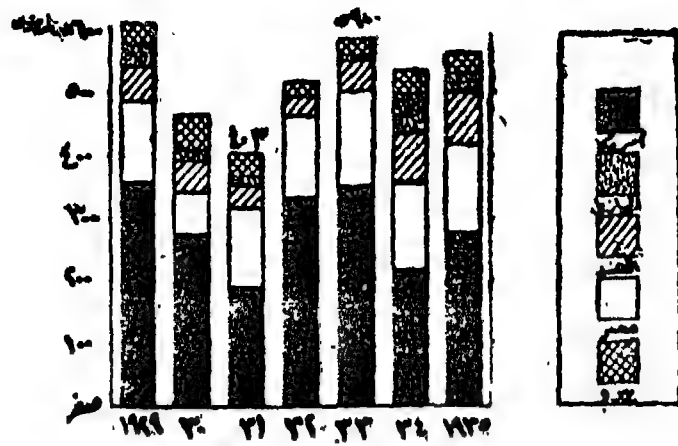


شكل رقم (٣-١٢)
النسب المئوية للمشتغلين بكل من الحرف الرئيسية
في أحد المحافظات في الفترة من ١٩١٧ - ١٩٤٧
«طرق الرسوم البيانية المجمعة النسبية»

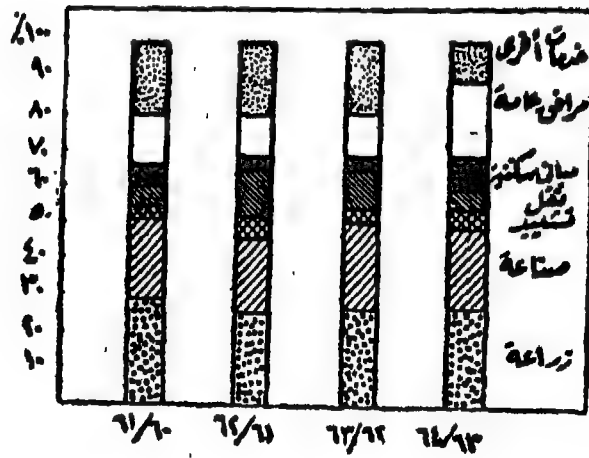
ويمكن رسم هذا النوع من الرسوم على أساس النسب المشوية هي بطبيعة الحال تكون الأنسب والأحسن . ويتم ذلك بتقسيم الظاهرة إلى أجزائها المختلفة بشرط أن تكون بنفس الترتيب لكل فترة زمنية، ثم نصل بين نقط التقسيم بخطوط وتظل المساحات المحصورة بين هذه الخطوط وبذلك يمكن معرفة عما إذا كانت نسبة أى قسم من الظاهرة قد هبطت أو زادت فى نفس الوقت بالنسبة إلى باقى التقسيمات الفرعية الأخرى للظاهرة (شكل رقم ٣-١٢).

٢ - الأعمدة البيانية المركبة Compound- Bar graphs

أما طريقة الأعمدة البيانية المركبة فهي عبارة عن أعمدة ذات عرض متساوى ومقسمة إلى أقسام داخلية تمثل فى مجموعها المجموع الكلى للظاهرة، وذلك بدلا من رسم عدة رسوم بيانية كل رسم منها يمثل جزءاً أو قسماً من المجموع الكلى للظاهرة. وفى هذه الحالة فإنه يمكن مقارنة كميات هذه الظاهرة من ناحية الكميات المطلقة شكل رقم ٣-١٣) - كما أنه يمكن مقارنة كميات هذه الظاهرة من الناحية النسبية وذلك بتحويل كمية كل قسم فرعى منها إلى نسبة مئوية. وتسمى الأعمدة البيانية فى هذه الحالة باسم الأعمدة المركبة النسبية (شكل رقم ٣-١٤). وفى هذه الحالة لا يمكن مقارنة كل عمود (مستطيل) بآخر ولكن يمكن مقارنة الجزئيات (التفاصيل) من كل عمود بالجزء الذى يناظره فى العمود الآخر، وذلك بمعرفة الفرق بين نسبتيهما بالنسبة للمجموع الكلى. ويجب فى هذه الحالة أن يصاحب الرسم البيانى للأعمدة المركبة النسبية. رسم بيانى آخر تكون أرقامه مطلقة حتى يمكن معرفة التغير بين المجموع الكلى لكل ظاهرة وأخرى.



شكل رقم (٣-١٣)
الواردات البريطانية من الأقطان الخام
(طريقة الأعمدة المركبة المطلقة)



شكل رقم (٣-١٤)
التطور النسبي لمكونات الدخل القومي لمصر
في الفترة ٦٠ - ١٩٦٤ (طريقة الأعمدة المركبة النسبية)

٣- الأهرامات البيانية Pyramid graphs

تستخدم الأهرامات البيانية كأحد طرق التمثيل البياني للبيانات الديموجرافية وبصفة خاصة لبيانات التركيب النوعي والعمرى للسكان، حيث يجمع الهرم البياني نسب كل من الذكور والإناث إلى العدد الكلى للسكان فى الفئات العمرية المختلفة. والهرم البياني عبارة عن أعمدة بيانية أفقية ترسم على محورين أفقيين أحدهما يمثل أعداد (أو نسب) السكان الذكور والآخر يمثل أعداد (أو نسب) السكان الإناث، أما المحور الرأسى للهرم فهو يمثل فئات العمر لكل من النوعين من السكان ويجب أن تقسم المحاور الأفقية بنفس المقياس سواء الذكور أو الإناث.

ومن المفيد استخدام هذا الأسلوب من التمثيل البياني إذا أريد معرفة الخصائص أو تشخيص الاتجاهات للمجتمعات السكانية، وكذلك عند إجراء المقارنات بين حالة السكان لأكثر من إقليم أو دولة لإظهار الصفات العامة للسكان باستخدام بيانات التعدادات السكانية. وهناك عدة أنواع من أهرامات السكان البيانية نوجز طريقة أنشاء كل منها فيما يلى:

أ- الهرم السكانى البسيط:

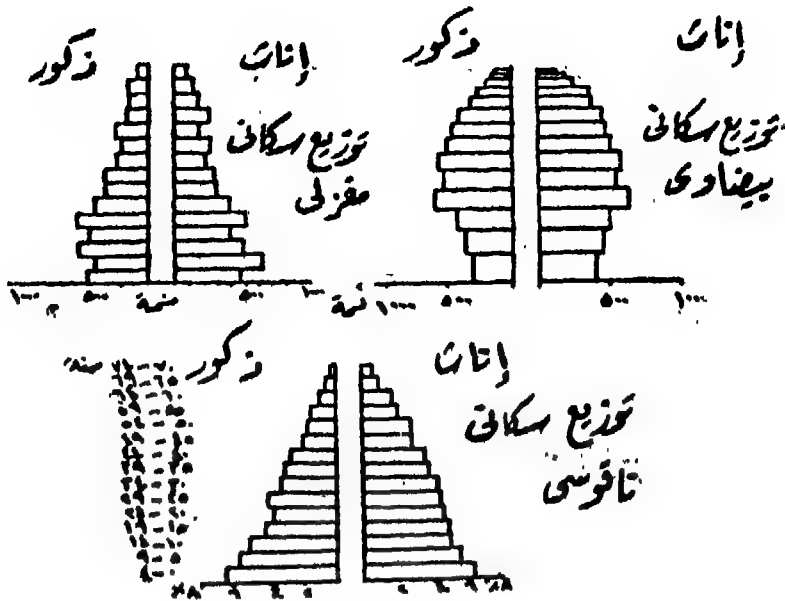
وتقوم فكرة إنشائه على الأساس السابق شرحة لإنشاء ورسم الهرم السكانى. ويستخدم هذا النوع من الأهرامات لبيان الصفات العامة لسكان دولة أو إقليم معين. ومن المعروف لدى علماء الديموجرافيا أن لكل دولة هزم سكانى يميز تراكيبها السكانى من حيث النوع والعصر لتعداد معين. وبناء على ذلك فإن أشكال الأهرامات السكانية ستختلف باختلاف التركيب النوعى والعمرى للسكان بين البلاد المختلفة وهذا الاختلاف هو الذى يبرز المميزات ويؤكد الاتجاهات السكانية بين البلاد المختلفة. وهذا الاختلاف هو الذى يبرز المميزات ويؤكد الاتجاهات السكانية التى بالتالى تعطى صورة واضحة عن التركيب العمرى والنوعى للمجتمعات السكانية لهذه الدول (شكل رقم: ٣-١٥) فمثلاً إذا كان الهرم السكانى يأخذ الشكل المغزلى المقلوب فإن ذلك يدل على أن المجتمع الذى يمثلته يتميز بتعادل معدلات المواليد والوفيات فيه. وإذا كان الهرم السكانى يأخذ

الشكل البيضاوى من أعلى (أى فى الفئات ذات الأعمار الكبيرة) وليس من عند القاعدة (أى فى الفئات ذات الأعمار الكبيرة) وليس من عند القاعدة (أى فى الفئات ذات الأعمار الكبيرة) وليس من عند القاعدة (أى فى الفئات ذات الأعمار الصغيرة) فانه يدل على أن المجتمع الذى يمثلته مجتمعنا مسنا. ويستنتج ذلك من انخفاض نسبة الأطفال (ذكور وإناث) وزيادة نسبة المسنين (الذكور والإناث). أما إذا كان الهرم السكانى يتخذ شكلا قريبا من شكل الناقوس (الجرس) حيث تكون قاعدته عريضة ومحدب بلطف فى قمته فإن ذلك يدل على ارتفاع معدلات الخصوبة.

ب- الاهرامات السكانية المركبة Compound Pyramids

وتقوم فكرة هذا النوع من الاهرامات السكانية على أساس تمثيل التركيب النوعى أو العمرى للسكان بأعمدة طول كل عمود يتناسب مع العدد الكلى للسكان لكل تعداد من التعدادات وبعد ذلك يقسم العمود (مثل طريقة الأعمدة المركبة) إلى أقسام الظاهرة الفرعية، كأن يقسم مثلا إلى سكان الريف (الزراع وغير الزراع) وسكان الحضر لكل تعداد. وفى الشكل رقم (٣-١٦) يمكن ملاحظة أنه خلال الفترة من ١٧٩٠ إلى ١٨٨٠ كان هناك فئتين فقط (سكان الريف والمخلات العمرانية التى يتراوح عدد السكان بها من ٢٥٠٠ إلى ٨٠٠٠ نسمة، بالإضافة إلى فئة السكان أكثر من ٨٠٠٠ نسمة والتى يمكن أن نعتبرها بمثابة سكان الحضر. أما فى الفترة من ١٨٩٠ إلى ١٩١٠ فإننا نلاحظ ثلاث تقسيمات عندما أدخلت فئة السكان ٢٥٠٠ - ٨٠٠٠ نسمة كفئة مستقلة.

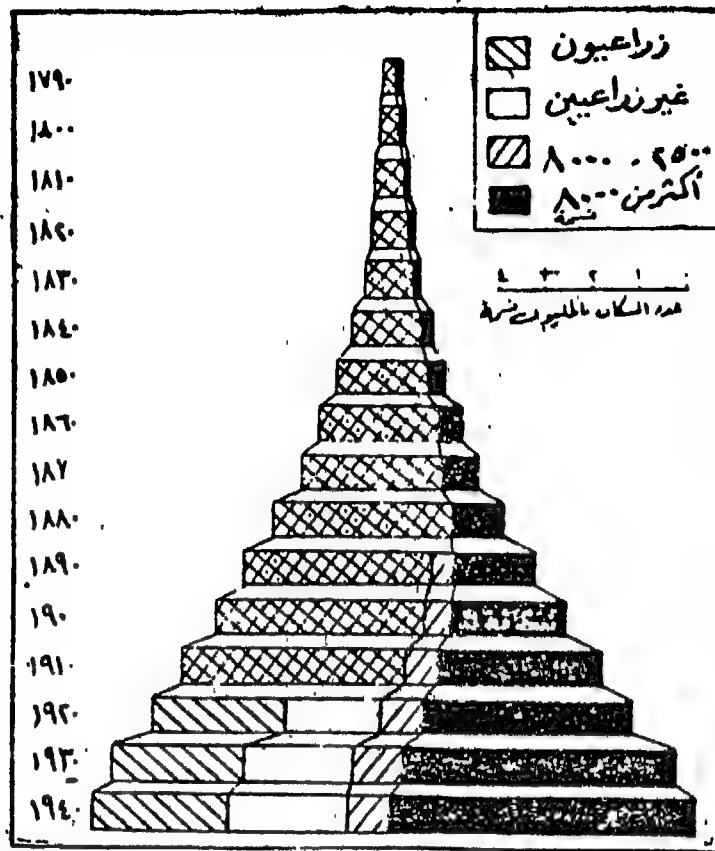
ويلاحظ على أعمدة الهرم من سنة ١٩٢٠ حتى ١٩٤٠ أنه قد ارتفع عدد تقسيماتها الداخلية لتضم فئتين لسكان المناطق الريفية من الزراع وغير الزراع، بالإضافة إلى فئتين لتضم فئتين أخريتين: فئة السكان الذين يتراوح عددهم من ٢٥٠٠ إلى ٨٠٠٠ نسمة، وأخرى للسكان أكثر من ٨٠٠٠ نسمة (سكان الحضر).



شكل رقم (٣-١٥)
أشكال الأهرامات السكانية البسيطة

ج- الأهرامات السكانية المنطبعة Superimposed Pyramids

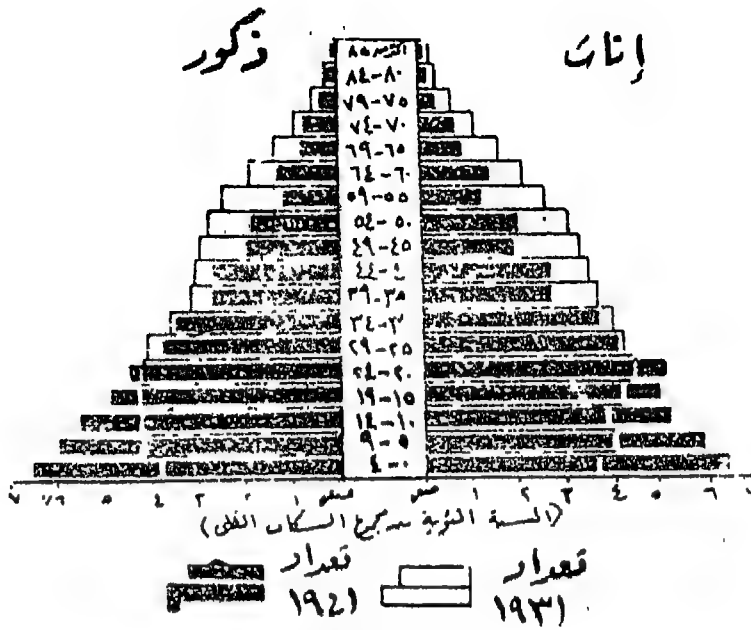
ويستخدم هذا النوع من الأهرامات لتمثيل بيانات التركيب النوعي والعمرى في مكان ما لعدة تعدادات مختلفة وذلك بقصد المقارنة بين عدد السكان في كل تعداد وآخر. كما يمكن استخدام هذه الطريقة لمقارنة حالة السكان من حيث التركيب العمرى والنوعى لمنطقتين للوقوف على مدى اختلاف توزيع السكان في أحدهما عن الأخرى.



شكل رقم (٣-١٦)

الهرم اليانسي المركب

وفي كل من الحالتين يمكن رسم هرم سكاني بسيط لأحد التعدادين أو لإحدى المنطقتين بالطريقة السابق ذكرها وإعطائه لونا أو ظلا معيناً، ثم يرسم بعد ذلك هرمًا بسيطاً آخر للسكان في التعداد الثاني أو في المنطقة الثانية بنفس مقاييس الرسم المستخدمة على الهرم السكاني الأول فيبدو وكأنه منطبقاً عليه (شكل رقم : ٣-١٧).



شكل رقم (٣-١٧)

الهرم السكاني المنطبع

٤- الدوائر البيانية المقسمة Divided Circles

يهدف رسم الدوائر البيانية المقسمة، التي تستخدم أساساً لمقارنة ظاهرتين أو أكثر، أو للمقارنة بين المكونات المختلفة لظاهرة واحدة بعضها البعض والمجموع الكلي خلال فترات زمنية متفاوتة، إلى إظهار التفاوت بين مجموع قيم الظواهر وبيان التغير بين المجموع الكلي لقيم الظاهرة الواحدة ومكوناتها. ويمكن أن نستفيد من استخدام هذه الدوائر في حالتين أساسيتين:

عندما يكون المجموع الكلي كبيراً نسبياً ولكنه يتمثل في مساحة محدودة جداً، عندما يكون المجموع الكلي كبيراً نسبياً ولكنه يتمثل في مساحة محدودة جداً، كما في حالة تمثيل عدد سكان المدن أو تمثيل إنتاج المصانع، أو عندما نريد تمثيل الكميات الكلية في منطقة أو إقليم أو دولة، كما في حالة تمثيل إنتاج البترول في البلاد العربية مثلاً.

ونظراً لأن مساحة الدائرة تتناسب مع مربع نصف قطرها (مساحة الدائرة = ط
نق^٢) فإنه يجب عند رسم مجموعة من الدوائر البيانية لعدة ظاهرات أخذ للجذر
التربيعي للمقيم الكلية التي تمثل كل ظاهرة ليكون بمثابة نصف قطر لكل دائرة
على حدة. أما إذا كان التمثيل البياني لظاهرة واحدة فقط فعلينا أن نختار طول
نصف القطر المناسب والذي يعطى مساحة لدائرة تتلائم مع مساحة ورقة الرسم
المراد تمثيل الظاهرة عليها.

ولتمثيل البيانات الموضحة في الجدول رقم (٣-٢) بطريقة الدوائر البيانية
تجرى الخطوات الآتية:

أ - يجمع الإنتاج في كل سنة حتى نحصل على المجموع الكلى (السوى)
للإنتاج في كل فترة زمنية.

ب - نستخرج الجذر التربيعي لمجموع الإنتاج في كل سنة على حدة، ويكون
الناجى ممثلاً لطول لنصف القطر (نق) الذى يريد أن نعرفه لكى ترسم
الدوائر التى تمثل الإنتاج، والجذور التربيعية للشال هى ١١,٦، ١١,٣،
١٢.

جدول رقم (٣-٢)

إنتاج مناطق الصيد بجمهورية مصر فى الفترة ٦١ - ١٩٦٤ (بالألف طن)

١٩٩٦٤/٦٣	١٩٩٣٤/٦٢	١٩٩٢/٦١	السنوات الإنتاج
٧٠	٦٥	٦٥	إنتاج البحار
٥٦	٥٤	٤٧	إنتاج البحيرات
١٨	١٦	١٥	إنتاج النيل
١٤٤	١٣٥	١٢٧	المجموع

ج- نختار قيمة قياسية أساسية سواء بالسنتيمتر أو المليمتر، يمكن على أساسها أن نحول أعداد الجذور التربيعية الناتجة لدينا إلى أطوال متناسبة تمثل مباشرة أنصاف أقطار الدوائر. وفي العادة تعطى هذه القيمة الأساسية لأصغر جذر تربيعي.

د- لمعرفة أنصاف أقطار الدوائر، هناك عدة طرق هامة تؤدي كلها إلى نتيجة واحدة ولن هنا تختلف في العمليات الحسابية. وسنختار من هذه الطرق طريقتين مألوفتين هما: طريقة التناسب الحسابي (طريقة المقص). ويمكن أن نطبقها على المثال السابق. فمثلاً إذا اخترنا الطول ١٦ ملليمتر كقيمة أساسية للجذر التربيعي ١١,٣ فإن:

$$(١) \quad ١١,٣ = ١٦ \text{ ملليمتر (نصف قطر الدائرة الأولى)}$$

$$١١,٦ = \text{س ملليمتر (نصف قطر الدائرة الثانية)}$$

$$\text{س} = \frac{١١٦ \times ١١,٦}{١١,٣} = ١٦,٤ \text{ ملليمتر}$$

$$(٢) \quad ١١,٣ = ١٦ \text{ ملليمتر (نصف قطر الدائرة الأولى)}$$

$$١٢ = \text{س ملليمتر (نصف قطر الدائرة الثالثة)}$$

$$\text{س} = \frac{١٢ \times ١٦}{١١,٣} = ١٧ \text{ ملليمتر}$$

والقيمة الأساسية التي اخترناها يعتمد اختيارها على مساحة لوحة الرسم ويجب عند اختيارها أن تتوافق مساحة الدوائر مع أبعاد مسطح لوحة الرسم بحيث لا تظهر أصغر دائرة صغيرة جداً وأكبر دائرة كبيرة جداً بالنسبة لمساحة لوحة الرسم.

أما الطريقة الأخرى فتعرف بطريقة الجذور التربيعية وهي طريقة سهلة ولا تتطلب كثيراً من الحساب وتتلخص في أن نقسم الجذور التربيعية على العدد ١٠ أو قوى هذا العدد الصحيح (١٠٠، ١٠٠٠، ١٠٠٠٠ ... الخ) وذلك طبعاً على

حسب المدى الذى توجد عليه الجذور التربيعية. ففى مثالنا السابق يمكن أن نقسم الجذور التربيعية كلها على العدد ١٠ ويكون تمييز الناتج بالسنتيمتر، وعلى هذا الأساس نجد أن:

$$\text{نصف قطر الدائرة الأولى} = ١١,٣ \div ١٠ = ١,١٣ \text{ سنتيمتراً}$$

$$\text{نصف قطر الدائرة الثانية} = ١١,٦ \div ١٠ = ١,١٦ \text{ سنتيمتراً}$$

$$\text{نصف قطر الدائرة الثالثة} = ١٢ \div ١٠ = ١,٢ \text{ سنتيمتراً}$$

وكما هو واضح فإن الأطوال التى نتجت بهذه الطريقة هى أطوال صغيرة وبالتالي ستكون مساحات دوائرها صغيرة. وفى مثل هذه الحالة يجب أن نكبر الأطوال الناتجة، وذلك بضربها كلها فى أى رقم نختاره، بحيث تظهر الدوائر بعد رسمها ملائمة لأبعاد لوحة الرسم. فإذا كان هذا الرقم الذى اخترناه هو ١,٥ مثلاً فسوف يصبح:

$$\text{طول نصف قطر الدائرة الأولى} = ١,١٣ \times ١,٥ = ١,٦٩٥ \text{ سنتيمتراً (١,٧٠ سم تقريباً)}$$

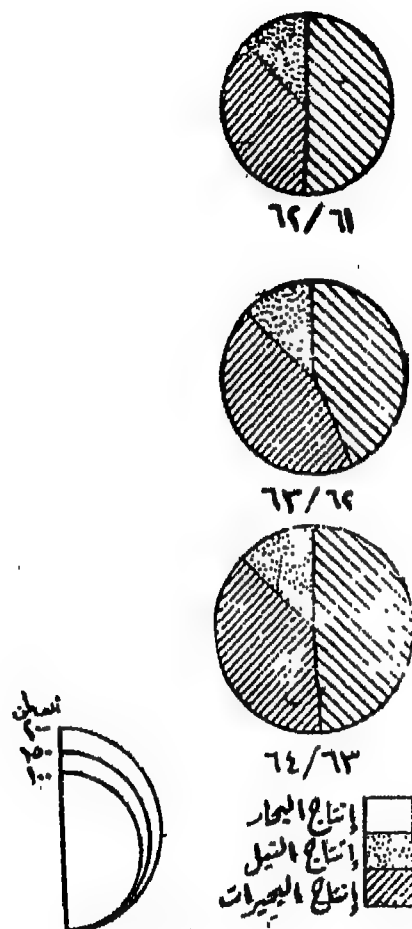
$$\text{طول نصف قطر الدائرة الثانية} = ١,١٦ \times ١,٥ = ١,٧٤٠ \text{ سنتيمتراً (١,٧٥ سم تقريباً)}$$

$$\text{طول نصف قطر الدائرة الثالثة} = ١,٢٠ \times ١,٥ = ١,٨٠ \text{ سنتيمتراً}$$

وأنصاف الأقطار الأخيرة التى رسمت على أساسها الدوائر فى الشكل رقم (٣-١٨) وسواء استخدمنا أى من الطريقتين السابقتين لمعرفة أطوال نصف قطر الدوائر فيجب أن لا نكتب عليها أية أعداد للكميات الحقيقية التى تمثلها الدوائر بنفس طريقة رسم الدوائر السابق شرحها ومنه يمكن أن نقيس قطر أى دائرة مرسومة، وليس من الضرورى أن يمثل هذا المقياس الدوائر المرسومة نفسها وإنما يمثل مقياساً لدوائر كمياتها ذات أرقام صحيحة بحيث تكون قريبة من الكميات الحقيقية التى تم تمثيلها بيانياً. وفى المثال السابق فإن هذه الكميات تكون ١٠٠، ١٥٠، ٢٠٠ (شكل رقم: ٣-١٨).

ويمكن تقسيم الدائرة إلى أقسام داخلية للمقارنة بين أجزاء الظاهرة. وفى هذه الحالة تحول الأرقام المطلقة إلى أرقام نسبية عن طريق قسمة رقم كل جزء على المجموع الكلى للظاهرة وضربه فى ١٠٠، ثم ضرب الناتج فى ٣,٦ فنحصل على

الزاوية المركزية التى تمثل هذا الجزء. وفى المثال السابق تقسم الدائرة الأولى (٦٢/٦١) إلى ثلاثة أقسام بنسبة إنتاج كل من البحار والبحيرات والنيل كمايلي:



شكل رقم (٣-١٨)

إنتاج مناطق الصيد فى مصر ٦١ - ١٩٦٤ (طريقة الدوائر المقسمة)

$$\% ٥١,٢ = ١٠٠ \times \frac{٦٥}{١٢٧} = \text{نسبة إنتاج البحار}$$

$$\% ٣٧ = ١٠٠ \times \frac{٦٥}{١٢٧} = \text{نسبة إنتاج البحيرات}$$

$$\% ١١,٨ = ١٠٠ \times \frac{١٥}{١٢٧} = \text{نسبة إنتاج النيل}$$

وتكون الزاوية المركزية لكل منها هي:

$$١٨٤ = ٣,٦ \times ٥١,٢ = \text{إنتاج البحار}$$

$$٢٢٣ = ٣,٦ \times ٣٧ = \text{إنتاج البحيرات}$$

$$٤٢ = ٣,٦ \times ١١,٨ = \text{إنتاج النيل}$$

وبالمثل يمكن حساب النسبة المئوية والزاوية المركزية للإنتاج المصايد في السنتين الأخيرتين كما في الجدول التالي:

جدول رقم (٣-٣)

تحديد الزاوية المركزية في الدائرة الممثلة لإنتاج المصايد المصرية

٦٤/٦٣		٦٣/٦٢		٦٢/٦١		السنوات الإنتاج
الزاوية المركزية	%	الزاوية المركزية	%	الزاوية المركزية	%	
١٧٥	٤٨,٦	١٧٢,٨	٤٨	١٨٤	٥١,٢	إنتاج البحار
١٤٠	٣٨,٩	١٤٤,٠	٤٠	١٣٣	٣٧,٠	إنتاج البحيرات
٤٥	١٢,٥	٤٣,٢	١٢	٤٢	١١,٨	إنتاج النيل
٣٦٠	١٠٠	٣٦٠,٠	١٠٠	٣٦٠	١٠٠	المجموع

ولتسهيل المقارنة بين أجزاء الظاهرة خلال فترة السنوات الثلاث يجب أن نأخذ الضلع الرأسى المكون للربع الأول من الدائرة (الخط الواصل بين مركز الدائرة وبداية تقسيم الدائرة أو صفر التدريج) ونعتبره كخط أساس يبدأ منه قياس الزوايا المركزية بعد تجميعها تصاعدياً. (شكل رقم : ٣-١٨).

٥- الرسوم البيانية المثلثية:

تستخدم الرسوم البيانية المثلثية في تمثيل البيانات النسبية الخاصة بثلاث ظاهرات مختلفة، أو في تمثيل البيانات الأساسية الخاصة بثلاثة عناصر لظاهرة واحدة (مثل بيانات العمالة في المصانع، أنواع الحيوانات، أنواع المحاصيل، أنواع النباتات، عناصر التربة) وذلك لمعرفة النسبة الغالبة بين الظاهرات أو الصفة السائدة بين عناصر الظاهرة بوجه عام.

وتقوم فكرة هذه الرسوم على أساس رسم مثلث متساوي الأضلاع، يقسم كل ضلع منه إلى عشرة أقسام متساوية تستخدم كمقياس نسبي يبدأ من الصفر حتى ١٠٠٪ ويكون التقسيم في اتجاه عقرب الساعة. أو بمعنى آخر أن يكون رقم ١٠٠ على أحد الأضلاع هو رقم الصفر للضلع المجاور والعكس مع رقم الصفر فيكون رقم ١٠٠٪ للضلع المجاور .. وهكذا. وبعد ذلك فصل بين كل رقم على أحد الأضلاع والرقم على الضلع المجاور ليكون مجموع هذين الرقمين ١٠٠٪ وبذلك نحصل على مجموعة من المثلثات الداخلية كل منها يشابه المقياس الكبير، ومنها نجد أن مجموع النسب لثلاثة عناصر إذ أضيفت لبعضها لحصلنا على الرقم ١٠٠٪ الذي يمكن تمثيله على الرسم البياني المثلثي بنقطة واحدة فقط. وفي عملية توقيع مكان هذه النقطة نبحث أولاً عن القيمة المراد تمثيلها على أحد الأضلاع الذي يمثل إحدى العناصر، ونمد منها خطاً يلتقى مع الخط الذي يمد من مكان القيمة الثانية على أحد الضلعين الآخرين. وسيلقى حتماً كل من الخطين مع الخط الواصل عن مكان القيمة الثالثة على الضلع الثالث. ومكان تلاقي الخطوط الثلاثة هو موقع النقطة التي ستجمع قيم الثلاث ظاهرات أو الثلاثة عناصر في موضع واحد على الرسم.

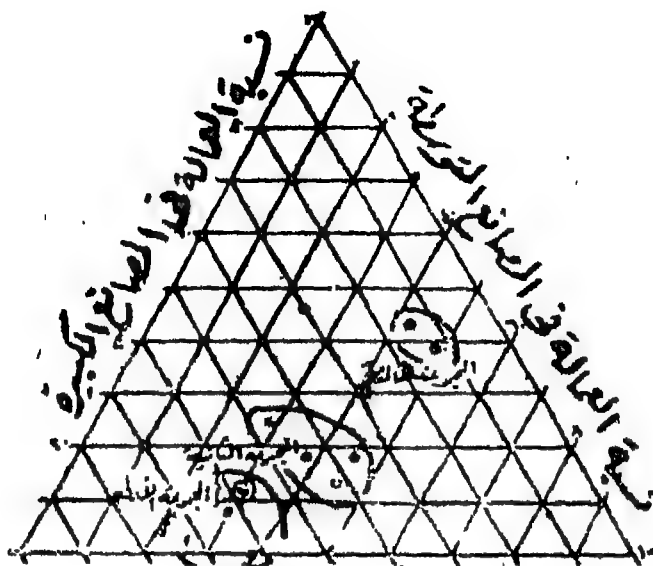
ومن الأمثلة الشائعة لاستخدام هذه الرسوم ما يختص بتحليل عينات التربة فمثلاً إذا كانت لدينا مجموعة من عينات التربة وكان تحليلها على أساس النسب المثوية للعناصر الثلاثة الرئيسية التي تتألف منها (الرمل، الغرين، الصلصال) وكان المطلوب تمثيلها بيانياً لمعرفة الصفة الغالبة للتربة بوجه عام، كان من الممكن

عندئذ استخدام هذا النوع من الرسوم البيانية. وبالمثل يمكن استخدامه لبيان الحالة العامة لثلاث من أنواع الصناعات في مجموعة من المدن.

ولهذه الرسوم البيانية أهمية خاصة إذ أنه يمكن أن تتخذ كأساس لوضع تصنيف أو استخراج أنماط لعدد من الظواهر عن طريق تحديد بعض المساحات على الرسم والتي منها يمكن أن نعرف موقع القيمة الثلاثية. فإذا كان هذا الموقع بالقرب من أحد أركان (نقط) المثلث يعني أن قيمة أحد العناصر (أحدى الظواهر) لا بد أن تكون كبيرة جداً، بينما وقوع القيمة الثلاثية بالقرب من جوانب المثلث يشير إلى أن قيمة أحد العناصر (أحدى الظواهر) لا بد أن تكون صغيرة جداً.

وكمثال تطبيقي يمكن سرده لبيان استخدام هذا النوع من الرسوم البيانية استخدمت بيانات العمالة في ١٢ مصنعاً رئيسياً بالسويد. وقسمت هذه المصانع حسب حجم العمالة المدربة بها (١٠٠ - ٥٠٠ عامل) إلى ثلاث فئات هي: مصانع صغيرة، مصانع متوسط الحجم، ومصانع كبيرة. والشكل رقم (٣-١٩) يوضح تمثيل بيانات الفئات الثلاث من المصانع، وقته يمكن أن نرى أن الصناعات السويدية يمكن أن تصنف إلى ثلاث مجموعات (بدون الصناعات الهندسية والحديدية التي تقف كمجموعة بمفردها): المجموعة الأولى تشمل على التحجير، الطباعة، الأعمال الخشبية، الصناعات الغذائية والمشروبات. وصناعات هذه المجموعة يقوم بانتاجها نسبة كبيرة من العمالة المدربة في المصانع الصغيرة لا تقل عن ٥٥٪ من جملة العمالة في الأنواع الثلاثة من المصانع، بينما تقل نسبة العمالة المدربة في المصانع الكبيرة في مجال إنتاج صناعات هذه المجموعة. والمجموعة الثانية تشمل الصناعات الجلدية، الغاز، المياه، والكهرباء، الصناعات الكيماوية، والنسيج. وهي صناعات لا يتعادل التدريب في إنتاجها بين الأنواع الثلاثة من المصانع حيث نجد أن هناك نسبة عمالة مدربة صغيرة للمصانع الكبيرة بينما تتعادل تقريباً نسبة العمال التي تدربها المانع المتوسطة الحجم والمصانع الصغيرة التي تهتم بأنشطة المجموعة الصناعية الثانية. أما المجموعة الثالثة فتمثل الصناعات التعدينية وتصنيع الورق، وهذه تتساوى فيها تقريباً نسبة العمالة المدربة في المصانع الكبيرة والمتوسطة

الحجم، بينما تقل نسبة العمالة المدربة كثيراً لإنتاج صناعات هذه المجموعة في المصانع الصغيرة الحجم.

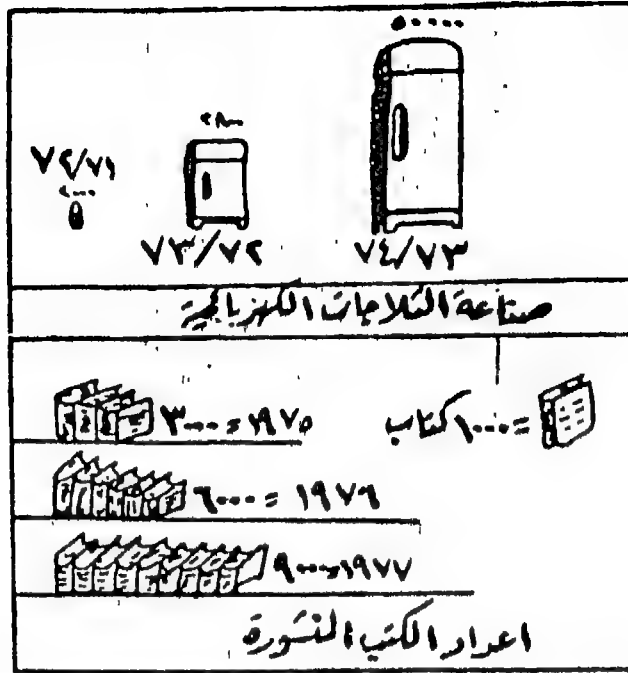


شكل رقم (٣-١٩)
تحديد فئات المصانع ونسبة العمالة بها
(طريقة الرسوم البيانية المثلثية)

ثالثاً: الطرق البيانية بالرسوم التصويرية Pictorial graphs

تعد طريقة الرسوم البيانية التصويرية من أحسن وسائل الإيضاح وأكثرها جاذبية في التعبير عن تغير وتطور كميات الظواهر. وتقوم فكرة هذه الطريقة على أساس إعطاء وحدات قياس (مفردات) الظاهرة أشكالاً تصويرية (شكل رقم ٣-٢٠). فمثلاً عدد السكان يمكن أن نمثله برسم عدد من الأشخاص يمثل الشخص الواحد عدد مليون نسمة، أو عدد الثلاثيات لأحد مصانع الثلاثيات يمكن تمثيله برسم عدد من الثلاثيات كل واحدة منها تمثل عشرة أو مائة ثلاثة. وكذلك عدد الكتب التي تصدرها إحدى دور النشر تصور بكتاب لكل

عدد معين من هذه الكتب (شكل رقم ٣-٢١)، كما أن عدد قراء إحدى الصحف اليومية يصور على أساس قارئ لكل عدد معين من الأعداد الصادرة.



شكل رقم (٣-٢١)
نموذج للرسم البيانية التصويرية

وفي كل الحالات السابقة يجب مراعاة الدقة في الرسم حتى يمكن الحصول على عدد من الرسوم التوضيحية أو الرموز تشابه تماماً الظاهرة المراد تمثيلها وعند تحديد عدد الوحدات من الظاهرة والتي يمثلها الشكل المختار فانه يجب تحديد هذا العدد في ضوء أكبر وأصغر رقم في البيانات وكذلك في ضوء مساحة اللوحة المخصصة للرسم. ويمكن كذلك تكبير أو تصغير الوحدة التي تمثل الظاهرة ويراعى أن يكون هذا التكبير والتصغير على أساس مقياس الرسم لعدد المرات التي تساويها الوحدة الصغيرة. أما كسور القيم فيمكن تمثيلها برموز أو أشكال غير كاملة.



شكل رقم (٣-٢٢)

أعداد قراء الصحف اليومية «طريقة الرسوم التصويرية»

ويعاب على هذه الطريقة رغم أنها تشد أنظار الشخص العادى فى التعرف على طبيعة الظاهرة من حيث تطورها وتغيرها، إلا أنها لا تعطى فكرة دقيقة عن قيم الظاهرة حيث أننا قد نضطر أحياناً إلى تقريب القيم إلى أقرب عشرة أو مائة أو ألف حتى يمكن التخلص من الكسور الصغيرة والتي لا يمكن أن نعطي لها شكلاً أو رمزاً. فمثلاً القيمة ٦٠٠ كتاب من الصعب تمثيلها تصويرياً برمز أو شكل متكامل إذا كان الأخير يمثل ١٠٠٠ كتاب مثلاً. كذلك تلائم هذه الطريقة تمثيل البيانات التي تتميز بتباين مفردات قيمها. وأيضاً لا تعطى هذه الطريقة فكرة حقيقية واضحة عن حقيقة المقارنة بين هذه الوحدات.

العرض البياني للبيانات المبوبة:

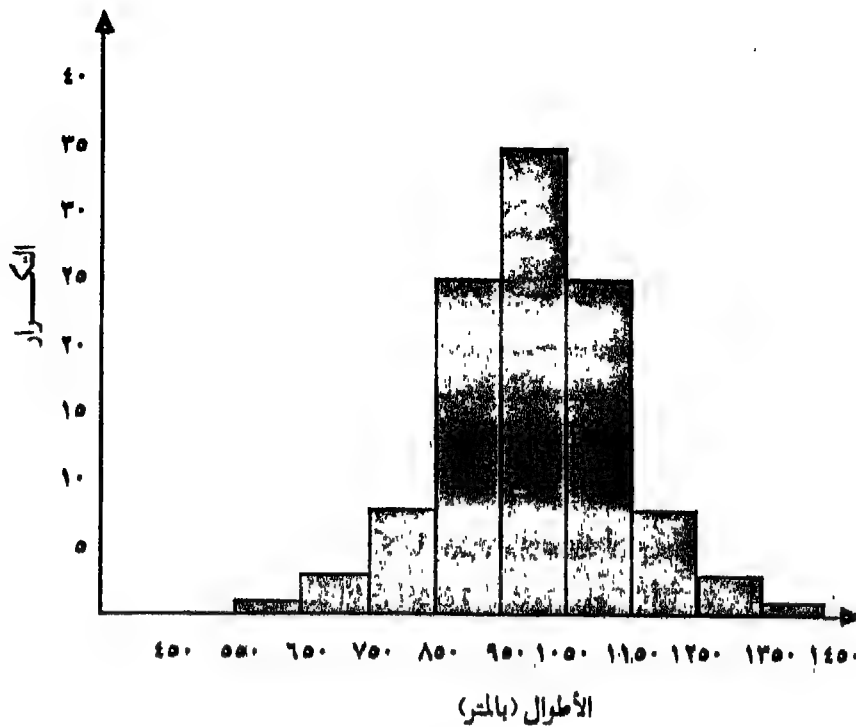
يقصد بالعرض البياني للبيانات المبوبة هو التمثيل التصويرى لبيانات جداول التوزيعات التكرارية. ويأخذ هذا العرض البياني أشكالاً مختلفة الهدف منها إعطاء صورة إجمالية عن خصائص البيانات وتوضيح الصفات المميزة للتوزيعات التكرارية.

وعموماً فإن بيانات جداول التوزيعات التكرارية تمثل بيانياً بعدة طرق مختلفة أهمها المدرج التكرارى Histogram ، المضلع التكرارى Polygon والمنحنى التكرارى Frequency Curve . .

المدرج التكرارى Frequency Histogram

يتكون المدرج التكرارى من مجموعة من المستطيلات المتلاصقة التى تكون قواعدها على المحور الأفقى «محور الستينات» متساوية وطول كل قاعدة منها يساوى طول فترة الفئة، كما يتناسب ارتفاع كل مستطيل مع التكرار المناظر لكل فئة على المحور الرأسى. ويراعى عند رسم المدرج التكرارى إختيار مقياس رسم ملائم لكل محور، بحيث تمثل جميع الفئات على المحور الأفقى كما يسمح المحور الرأسى بتمثيل جميع التكرارات. كما يجب ملاحظة أن يبدأ المحور الرأسى دائماً بالصفر حتى يمكن مقارنة التكرارات المختلفة بعضها ببعض. أما المحور الأفقى فليس من الضرورى أن يبدأ بالصفر، إذ أن ذلك سوف يحتاج منا إلى محور طويل جداً لتمثيل الفئات.

وإذا أردنا تمثيل التوزيع التكرارى لأطوال مائة رافد نهري (جدول رقم ٢-١١) برسم مدرج تكرارى له، نلاحظ أن التوزيع منتظم وينقسم إلى ٩ فئات متساوية، ولذلك نقوم بتقسيم المحور الأفقى إلى ٩ أقسام متساوية (ويستحسن أن نزيد عليها قسماً أى تصبح ١٠ أقسام). ويفضل أن نبدأ تقسيم المحور الأفقى بفئة أصغر من أقل فئة بالجدول أى نبدأ بتقسيم المحور الأفقى بالعدد ٤٥٠. ولما كان أكبر التكرارات هو ٣٣ فإننا نقوم بتقسيم المحور الرأسى إلى أقسام متساوية بمقياس رسم مناسب بحيث يسمح بظهور أكبر التكرارات بالجدول بحيث يبدأ التقسيم على المحور بالصفر. ثم نرسم على كل فئة من الفئات مستطيلاً تكون قاعدته مساوية لطول فترة الفئة وارتفاعه يساوى تكرار الفئة المقام فوقها هذا المستطيل. وبذلك نحصل على المدرج التكرارى الموضح فى شكل رقم (٣-٢٣)، والذي فيه مساحة المستطيلات تتناسب مع التكرارات (لأن أطوال قواعد المستطيلات جميعها متساوية ولأن مساحة كل مستطيل تساوى طول القاعدة فى الارتفاع).



شكل رقم (٣-٢٣)
المدرج التكرارى لأطوال ١٠٠ رافد نهري (بالمتر)

وإذا كانت الفئات كلها لها نفس الطول، فإنه من المعتاد أن ترسم الارتفاعات بأطوال مساوية لتكرارات الفئات. أما إذا كانت الفئات غير متساوية الطول فإن هذه الأطوال يجب أن تعدل قبل رسم المدرج التكرارى. وذلك بقسمة تكرار كل فئة على طول هذه الفئة وبذلك نحصل على التكرار المعدل، ثم ننشئ المدرج التكرارى برسم عدد من المستطيلات قاعدتها تمثل أطوال الفئات «غير المتساوية» على المحور الأفقى وارتفاعاتها تمثل التكرارات المعدلة. وفى هذه الحالة تتناسب مساحة كل مستطيل مع التكرارات المعدلة المناظرة لكل فئة.

وما تجدر الإشارة إليه أن المدرج التكرارى يصلح لتمثيل بيانات المتغيرات

المستمرة «المتصلة» ولا يصلح لتمثيل بيانات المتغيرات غير المستمرة «الوثابة»، إذ يجب أن تكون حدود فئات التوزيع المرسوم له الهيستوجرام معروفة تماماً. كما أن المدرج التكرارى يعبر عن اتجاهات لتوزيع التكرارى عن طريق إبراز الصفات العامة للتوزيع.

المضلع التكرارى Frequency Polygon

يمكن اعتبار إنشاء المدرج التكرارى بالطريقة التى ذكرناها سابقاً خطوة من خطوات رسم المضلع التكرارى للتوزيع التكرارى. وتعتمد فكرة إنشاء المضلع التكرارى على فكرة التمثيل البيانى من خلال الخط البيانى حيث تبين فقط تمثيل تكرار الفئة المقابلة لمركز الفئة. ويرسم المضلع التكرارى بإيصال نقط تنصيف المستطيلات المكونة للمدرج التكرارى بمجموعة من الأضلاع يقع كل ضلع منها بين مركز فئة ما ومركز الفئة التالية مباشرة. وحتى يمكن قفل شكل المضلع التكرارى وتحديد مساحته فإننا نفترض وجود فئتين إحداهما قبل الفئة الأولى والأخرى بعد الفئة الأخيرة والتكرار المناظر لكل منهما يساوى صفراً، ثم يتم توصيلهما بأضلاع مع مركزى الفئة الأولى والأخيرة فتحصل بذلك على شكل كامل للمضلع التكرارى كما فى شكل رقم (٣-٢٤).

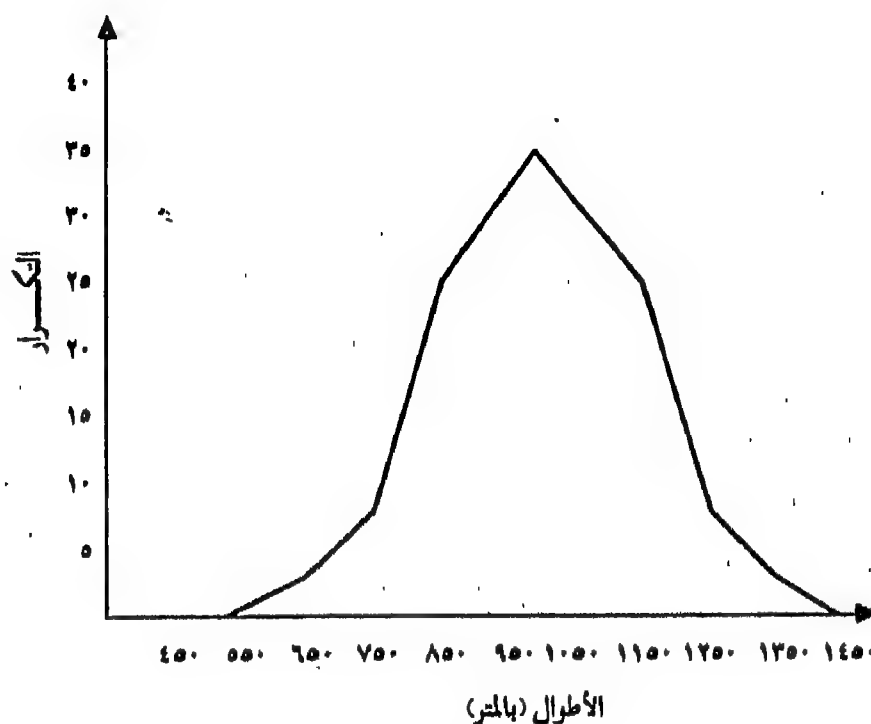
وفكرة إستخدام مركز الفئة أو منتصفها فى رسم المضلع التكرارى يعتمد على افتراض تركيز كل القيم عند متوسطها الحسابى حيث أن مركز الفئة يساوى:

$$\frac{\text{الحد الأدنى للفئة} + \text{الحد الأعلى للفئة}}{2}$$

٢

ويتميز المضلع التكرارى بأن المساحة المحصورة تحت المضلع هى نفسها المساحة المحصورة تحت المدرج التكرارى، ولكنه يكون أكثر دقة من المدرج التكرارى من حيث اعطائه صورة أكثر واقعية لاتجاهات وخصائص التوزيع. كما يعد المضلع التكرارى من أنسب الطرق البيانية لتمثيل أكثر من توزيع واحد من التوزيعات التكرارية مما يسهل إجراء المقارنة بينها. ويجب ملاحظة أنه لا يمكن إيجاد تكرار أية قيمة تقع بين مركزى فئتين من المضلع التكرارى وذلك لأن المضلع التكرارى

يمثل تكرار الفئات لا تكرار القيم الفردية كما هو واضح في الشكل رقم (٢٤-٣).



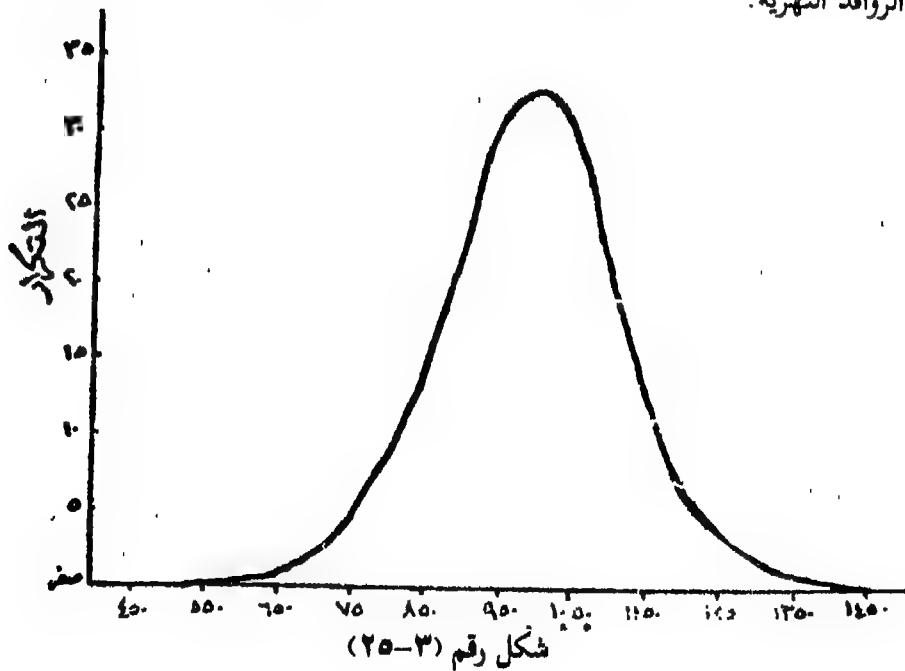
شكل رقم (٢٤-٢)
المضلع التكرارى لأطوال ١٠٠ رافد نهري «بالمتر»

المنحنى التكرارى: Frequency Curve

يمكن تمثيل التوزيعات التكرارية بطريقة أخرى، غير الطريقتين السابقتين، تظهر في شكل هندسى واضح وذلك برسم المنحنى التكرارى للتوزيع والذى نحصل عليه من خلال رسم المضلع التكرارى أولاً ثم تمهيد خطوط المضلع المنكسرة. وتقوم فكرة المنحنى التكرارى على أساس اعتبار أن البيانات تمثل عينة مسحوة من مجتمع أكبر. وبما أن هناك عدداً كبيراً من المفردات في المجتمع فإنه من الممكن من الناحية النظرية (للبيانات المستمرة) إختيار فترة الفئة صغيرة جداً

ويظل لدينا عدد ملموس من المفردات (تكرارات) في داخل كل فئة. وبهذا فإنه من المتوقع أن يتكون المضلع التكرارى للمجتمعات الكبيرة من عدد كبير من الخطوط الصغيرة المنكسرة والتي يمكن تقريها بمنحنى. وتزيد درجة الدقة في التقريب بزيادة حجم العينة، ولهذا السبب فإن المنحنى التكرارى يسمى أحياناً المضلع التكرارى الممهد.

ولرسم المنحنى التكرارى نعين مراكز الفئات على حسب التكرارات المناظرة لها، ونوقع نقط المضلع التكرارى ونمهد الخطوط المنكسرة بين هذه النقط باليد بحيث يمر المنحنى بأغلبية رؤوس المضلع التكرارى. (يوجد أيضاً طرقاً رياضية لتكوين المنحنى التكرارى تجعل المساحة تحت المنحنى مساوية للمساحة تحت المضلع التكرارى). والشكل رقم (٣-٢٥) يبين المنحنى التكرارى لتوزيع أطوال الروافد النهرية.



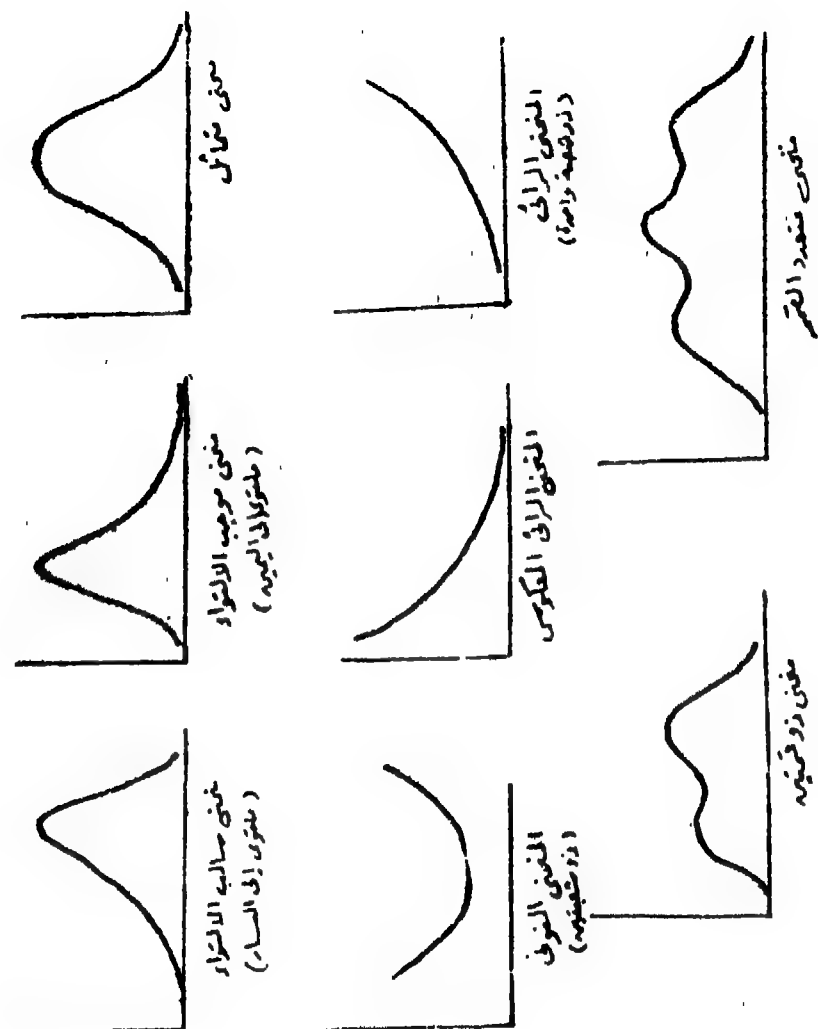
المنحنى التكرارى لأطول ١٠٠ رافد نهري (بالمتر)

ونظراً لأن المنحنى التكرارى يمكن رسمه من نقط المضلع التكرارى، لذا فإن شكل المنحنى يتوقف على توزيع البيانات ومن المنطقي أن نتوقع وجود عدد كبير من الأشكال المختلفة للمنحنيات التكرارية، إلا أنه يمكن حصر أشكال المنحنيات التى تقابلنا عادة فى التحليل الاحصائى للبيانات، والتى يمكن تقسيمها إلى نوعين رئيسيين:

النوع الأول: هو المنحنيات المتماثلة Symmetrical Curves. وهى المنحنيات التى تتماثل حول خط رأسى يقسم المنحنى إلى قسمين متطابقين أو بمعنى آخر هى المنحنيات التى تتماثل فيها التكرارات على جانبى أكبر تكرار. أما النوع الثانى: فهو المنحنيات غير المتماثلة Asymmetrical Curve وهى المنحنيات التى تظهر فيها صفة التماثل كأن تبدو فيها التكرارات متناقصة فى طرف عنه فى الطرف الآخر. ويأخذ كل نوع من هذين النوعين من المنحنيات أشكالاً مميزة كما هو كما هو موضح بالشكل رقم (٣-٢٦) ونعرضها فيما يلى:

١- المنحنى التكرارى المتماثل Normal Curve وهو أكثر المنحنيات شيوعاً ويأخذ هذا المنحنى الشكل الناقوسى Bell- shaped Curve الذى يتميز بأن المفردات المتساوية البعد عن مركز النهاية العظمى لها نفس التكرارات. ومن الأمثلة الهامة له المنحنى المعتدل وهو المنحنى الذى نتوقع الحصول عليه من دراسة كثير من الظواهر التى تتغير تبعاً لأسباب طبيعية مثل الطول والوزن..... الخ. ونظراً للأهمية الكبيرة للمنحنى المعتدل فى الدراسات الإحصائية لذا سندرسه بالتفصيل فيما بعد.

٢- المنحنيات التكرارية المئوية Skewed Culves وهى المنحنيات التى تتميز بأن أحد طرفيها يمتد أكثر من الآخر على جانبى مركز النهاية العظمى. فإذا كان الطرف الأيمن للمنحنى أطول من الطرف الأيسر فإن المنحنى يكون فى هذه الحالة ملتوياً إلى اليمين أو ملتوياً لتواء موجباً Positively Skewed- Curve بينما لو كان العكس صحيحاً فإن المنحنى يكون ملتوياً إلى اليسار أو ملتوياً لتواء سالباً Negativcly Skewcd- Curve.



شكل رقم (٣-٢٦) أشكال المنحنيات التكرارية

٣- المنحنيات ذات الشكل الرأسي أو الشكل الرأسي المعكوس. وفيها تقع نقطة النهاية العظمى عند أحد طرفي المنحنى. ويطلق على هذين النوعين من المنحنيات اسم المنحنيات الأسية Exponential الموجبة والسالبة على الترتيب. ومن أمثلة هذه المنحنيات منحنيات توزيع الثروة أو الملكية في المجتمع.

٤- المنحنى النوني، وهو المنحنى الذي يأخذ شكل حرف «U» U-shaped Curve. ويتميز هذا المنحنى بأن له نهاية عظمى (التكرارات الكبرى) عند كل من طرفيه بينما تكون النهاية الصغرى أو التكرارات الأقل في اتجاه مركز المنحنى. ويعتبر توزيع الوفيات للسكان حسب السن من أوضح أمثلة هذا الشكل من المنحنيات.

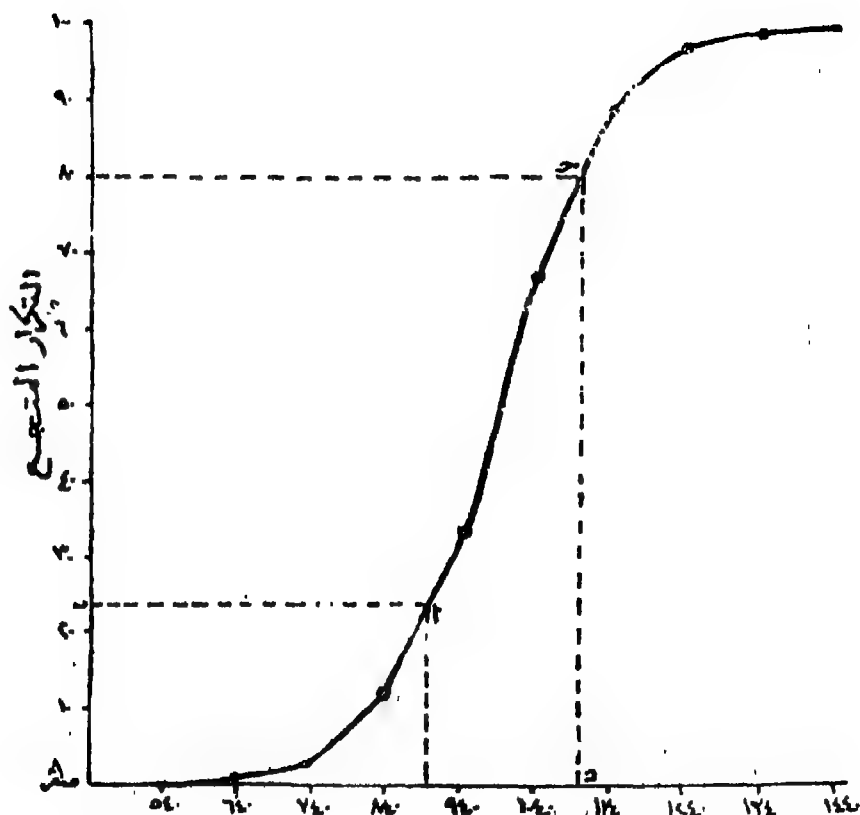
٥- المنحنى ذو القمتين Bimodal Curve، وهو المنحنى الذي يتميز بأن له نهايتان عظيمتان، والسبب في ذلك يرجع عادة إلى عدم تجانس العينة موضع الدراسة فقد تحتوي على مجموعتين مختلفتين ومتداخلتين من المفردات.

٦- المنحنى متعدد القمم Multi-mode، وهو المنحنى الذي له أكثر من نهايتين عظيمتين ويدل على عدم تجانس المفردات في العينة بمعنى أن تكون العينة مكونة من خليط من المجموعات.

المنحنى التكرارى المتجمع: Cumulative Frequency Curve (Ogive)

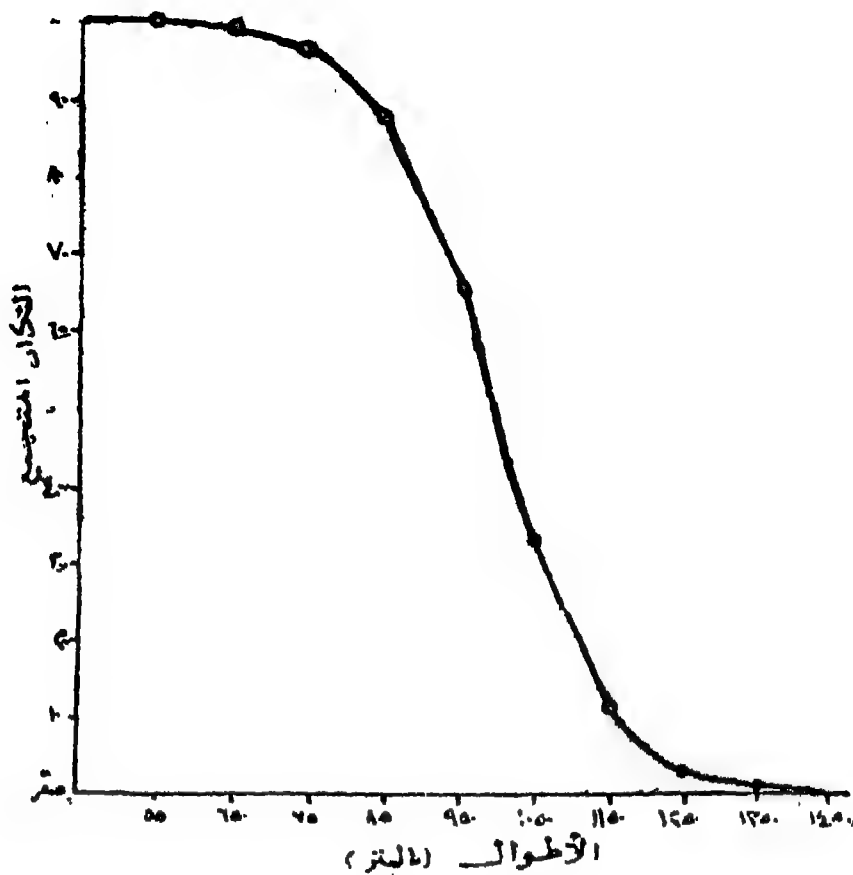
إلى جانب الأنواع السابقة من المنحنيات التكرارية هناك منحنيات بيانية تمثل التوزيعات التكرارية المتجمعة. وقد سبق أن أوضحنا كيفية عمل جداول التكرارات المتجمعة الصاعدة والهابطية وكذلك التكرارات المتجمعة النسبية. وتعرض جداول التوزيع التكرارى المتجمع بياناً على هيئة مصلع تكرارى متجمع أو منحنى تكرارى متجمع Ogive وهو إما من النوع الصاعد (أقل من) أو من النوع الهابط (أكبر من). ففي الحالة الأولى تمثل الحدود العليا للفئات على المحور الأفقى والتكرارات المتجمعة على المحور الرأسى، وتوقع القيم حسب إحداثياتها السينى والصادى بنقط فصل بينها بمنحنى ممدود فنحصل على المنحنى المتجمع الصاعد (شكل رقم ٢٧-٣) لأن التكرارات المتجمعة تكون فى تزايد. أما فى الحالة الثانية فتتمثل

الحدود الدنيا للفتات الأصلية على المحور الأفقى والتكرارات المتجمعة الهابطة على المحور الرأسى، ثم نصل بين النقط التى تمثل القيم فى الجدول بمنحنى ممهد فنحصل على المنحنى المتجمع الهابط، كما فى الشكل رقم (٢٨-٣).



شكل رقم (٢٧-٣)

المنحنى المتجمع الصاعد لعدد ١٠٠ رافد نهري بالمتر



شكل رقم (٣-٢٨)

المنحنى المتجمع الهايطة لعدد ١٠٠ رافد نهري وبالتر

وتتميز المنحنيات التكرارية المتجمعة (الصاعدة والهابطة) بأنه يمكن إستخدامها لإيجاد تكرار أى قيمة وسيطية تقع ما بين مركزي فئتين، وكذلك إيجاد الحد الأعلى لأى فئة يقع بداخلها تكرار معين. فمثلاً إذا أردنا معرفة عدد الروافد التى تقل أطوالها عن ٩٠٠ متر فإننا نقيم عموداً على المحور الأفقى عند القيمة ٩٠٠ حتى يقابل المنحنى المتجمع الصاعد فى نقطة (أ) ثم نرسم منها خطاً أفقياً موازياً للمحور الأفقى حتى يلاقى المحور الرأسى فى نقطة (ب) نحدد لنا عدد الروافد المطلوبة وهو ٢٤ (شكل رقم ٣-٢٧). ويحدث العكس إذا أردنا معرفة الحد الأعلى لأطوال ٨٠ رافداً فإننا نرسم خطاً أفقياً موازياً للمحور الأفقى عند ٨٠ حتى

يقابل المنحنى فى نقطة (ح) ثم نسقط منها عموداً على المحور الأفقى فيقطعه فى نقطة (د) تمثل الحد الأعلى لأطوالها وهو ١١٠٠ متراً من الرسم. أى أن هناك ٨٠ رافداً نهرياً أطوالها أقل من ١١٠٠ متراً.

وفى حالة إذا استخدمنا التكرارات المتجمعة النسبية لرسم منحنى تكرارى (صاعد أو هابط) بنفس الطريقة المتبعة فى رسم المنحنيات المتجمعة، فإن النتيجة تسمى بالمنحنى التكرارى المتجمع النسبى Cumulative percentage frequency curve

منحنى لورنز Lorenz Curve

يعد منحنى لورنز كأحد مؤشرات التفاوت Index of difference أو التركيز من أهم تطبيقات منحنى التوزيع المتجمع الصاعد النسبى. وحيث أن التركيز هنا يعنى سوء أو عدم عدالة التوزيع Inequality ، فإن الغرض من استخدام منحنى لورنز هو بيان درجة التفاوت فى توزيع الظاهرات ومقارنة عدالة توزيع لهذه الظاهرات من مكان لآخر أو من فترة زمنية إلى أخرى لنفس المكان. ومن أوضح نماذج البيانات الجغرافية التى يستخدم فى تحليلها منحنى لورنز بيانات توزيع ملكية الأراضى الزراعية وتوزيع العمالة على الصناعات المختلفة.

ويعتمد أسلوب منحنى لورنز فى التحليل على وجود توزيع نظرى (أمثل) Hypothetical or even distribution للظاهرة موضع الدراسة، ويتمثل هذا التوزيع بخط مستقيم يعطى أى نقطة عليه النسب المتساوية فى التوزيع. فمثلاً إذا كنا بصدد قياس علاقة مساحة الأراضى الزراعية بالملك فإنه من المفروض (نظرياً) أن نسبة ١٠٪ من الملاك يمتلكون ١٠٪ م الأراضى كما أن ٧٠٪ من الملاك يمتلكون ٧٠٪ من الأراضى.. وهكذا. كما يعتمد أسلوب منحنى لورنز أيضاً على وجود منحنى للتوزيع الفعلى للظاهرة. ولما كان التوزيع الفعلى لأى ظاهرة يختلف حتماً عن التوزيع النظرى (المثالى) لها، فإن العلاقة بين منحنى التوزيع الفعلى وخط التوزيع المتساوى تحدد درجة عدالة أو سوء توزيع هذه الظاهرة. ويستدل على ذلك من المساحة المحصورة بين المنحنى وخط التوزيع المتساوى

(مساحة التركيز أو سوء التوزيع)، فكلما قرب المنحنى من هذا الخط كلما صغرت المساحة بينهما وذل ذلك على قرب التوزيع من التوزيع المثالى أو اتخذ دليلاً على عدالة التوزيع، أما إذا بعد منحنى التوزيع الفعلى عن خط التوزيع المتساوى فإن المساحة المحصورة بينهما تزداد ويدل ذلك على بعد التوزيع عن التوزيع المثالى أو يتخذ دليلاً على سوء توزيع الظاهرة.

ولرسم منحنى لورنز لتوزيع الملكية الزراعية مثلاً (جدول رقم ١٣) تتبع الخطوات التالية:

١- تحول التكرارات الأصلية التى تمثل عدد الملاك والمساحة المملوكة إلى تكرارات متجمعة صاعدة.

٢- تحول التكرارات المتجمعة المطلقة إلى تكرارات متجمعة نسبية صاعدة وبالتالى نحصل على التوزيع الفعلى الذى يتم تعيينه على الرسم البيانى.

٣- نرسم مربعاً يمثل أحد محوريه الأفقيين التكرارات المتجمعة النسبية لعدد الملاك ويمثل أحد محوريه الرأسيين التكرارات المتجمعة النسبية لجملة المساحة، بحيث يبدأ مقياس كل محور بالصفر وينتهى برقم ١٠٠٪. توقع النقاط الممثلة للنسب على الرسم البيانى ونصل بينها بخط ممهد يطلق عليه منحنى التوزيع الفعلى (منحنى لورنز).

٤- نرسم الخط المستقيم الذى يصل بين نقطتى الصفر و ١٠٠ (قطر المربع) ليمثل التوزيع المثالى (النظرى). للتوزيع كما فى الشكل رقم (٣-٢٩).

جدول رقم (٣-٤)

توزيع الملكية الزراعية حسب فئات المساحة

وجملة المساحة المملوكة قبل صدور قانون الإصلاح الزراعى عام ١٩٥٢

التكرار المتجمع النسبى الصاعد		التكرار المتجمع الصاعد		جملة المساحة المملوكة (لأقرب ألف)	عدد الملاك (أقرب ألف)	فئات المساحة (الفدان)
% الجملة المساحة المملوكة	% لعدد الملاك	جملة المساحة المملوكة	% لعدد الملاك			
٣٥,٤	٩٤,٣	٢١٢٢	٢٦٤٢	٢١٢٢	٢٦٤٢	أقل من ٥
٤٤,٢	٩٧,١	٢٦٤٨	٢٧٢١	٥٢٦	٧٩	أقل من ٥ - ١٠
٥٤,٩	٩٨,٨	٣٢٨٦	٢٧٦٨	٦٣٨	٤٧	أقل من ١٠ - ٢٠
٥٦,٨	٩٩,٦	٣٩٤٠	٢٧٩٠	٦٥٤	٢٢	أقل من ٢٠ - ٥٠
٧٣,٠	٩٩,٨	٤٣٧٠	٢٧٩٦	٤٣٠	٦	أقل من ٥٠ - ١٠٠
٨٠,٣	٩٩,٩	٤٨٠٧	٢٧٩٩	٤٢٧	٣	أقل من ١٠٠ - ٢٠٠
١٠٠,٠	١٠٠,٠	٥٩٨٤	٢٨٠١	١١٧٧	٢	٢٠٠ فأكثر
				١١٨٤	٢٨٠١	المجموع

الباب الثانى مقاييس الوصف

مقدمة

الفصل الرابع: مقاييس النزعة المركزية

الفصل الخامس: مقاييس الانحراف (التشتت) والاختلاف

الفصل السادس: مؤشرات التركيز

الباب الثانى

مقاييس الوصف

مقدمة:

ذكرنا فى الباب السابق أن عملية تبويب وتنظيم البيانات الإحصائية عن طريق وضعها فى جداول أو تمثيلها بيانياً لا تتم بغرض عرض البيانات بصورة تلخص معاملها فحسب، بل أنها أيضاً تعد الخطوة الأولى على طريق التحليل الإحصائى لهذه البيانات. وحيث أن أسلوب الجدولة والتمثيل البيانى فى دراسة الظواهر يعتمد على دقة الأسلوب نفسه، فإن الخطوة التالية فى التحليل هى وصف البيانات بطريقة موضوعية غير متحيزة واستخلاص النتائج منها. ونلجأ فى وصف البيانات Data Description إلى حساب بعض أساليب القياس الكمية التى تعرف باسم «المقاييس الوضعية أو مقاييس الوصف الإحصائى». وهى إما مقاييس محسوبة من المجتمع وتسمى بمعالم (ثوابت) المجتمع Paramaters، أو مقاييس محسوبة من العينة وتعرف باسم إحصائيات العينة a statistic.

ونظراً لاختلاف الظواهر الجغرافية (طبيعية وبشرية) من حيث الخصائص والاتجاهات وبالتالى اختلاف نوعية البيانات وتنوع توزيعاتها التكرارية، فلقد تعددت أدوات القياس الكمية التى تهتم بتحديد هذه الخصائص والاتجاهات. فمثلاً قد تختلف التوزيعات فى القيمة المتوسطة التى تتركز حولها قيم المفردات، أو قد

أو في اتجاه وشكل تركيز قيم المتغيرات في أحد نواحي التوزيع وتعرف المقاييس الكمية في الحالة الأولى بمقاييس النزعة المركزية، وتعرف المقاييس في الحالة الثانية بمقاييس التشتت (الإختلاف)، بينما تعرف المقاييس في الحالة الثالثة بمقاييس أو مؤشرات التركيز. وسنعرض في هذا الباب أنواع مقاييس الوصف الكمية وطرق استخدامها وخصائصها ومجالات تطبيقاتها المتعددة في دراسة البيانات الاحصائية بأنواعها المختلفة، إلى جانب عرض مفصل لمزايا استخدام هذه المقاييس ومشاكل تطبيقاتها.

الفصل الرابع
مقاييس النزعة المركزية
Measures of Central Tendency

الفصل الرابع

مقاييس النزعة المركزية

المقصود بالنزعة المركزية هو نزعة المفردات للتركز حول قيمة متوسطة أو قيمة نموذجية تمثل مجموعة من البيانات. وحيث أن مثل هذه القيمة النموذجية تميل إلى الوقوع في المركز داخل مجموعة بيانات مرتبة حسب قيمها، فقد اتخذت كأساس للوصف الإحصائي لمعالم المجموعة التي تشكلها هذه البيانات. ويمكن أن نعرف صوراً عديدة لمقاييس النزعة المركزية وإن كان من أكثرها شيوعاً: المتوسط بأنواعه (المتوسط الحسابي، المتوسط الهندسي، المتوسط التوافقي)، الوسيط والمنوال وكل مقياس من هذه المقاييس له مميزاته وعيوبه وهنا يعتمد على البيانات والهدف من استخدامه. ويطلق في بعض الأحيان على المقاييس السابقة للنزعة المركزية اسم مقاييس المتوسطات Averages أو مقاييس الموضع Location.

أولاً: المتوسط Mean

يختلف تعريف المتوسط للبيانات تبعاً لنوع المقياس المستخدم، إلا أنه يعتبر أبسط مقاييس النزعة المركزية وأكثرها دقة وتداولاً. وتوجد عدة أسس لتحديد قيمة المتوسط لمجموعة من البيانات مما أدى إلى وجود عدد من مقاييس المتوسط أهمها: المتوسط الحسابي، المتوسط الهندسي والمتوسط التوافقي.

(١) المتوسط الحسابي Arithmetic Mean

يعتبر المتوسط الحسابي من أشهر مقاييس المتوسطات وأسهلها حساباً وأكثرها

استخداماً في عملية وصف البيانات. فهو يلخصها في قيمة محددة تتركز حولها أغلب المفردات، فتمثل لها الاتجاه عام وتعتبر بالتالي من أحد معالم ومميزات هذه البيانات. والمتوسط الحسابي مهم جداً في الدراسات الجغرافية لأنه يعطي فكرة عامة عن الظاهرة موضع الدراسة. كأن نعرف متوسط سقوط الأمطار في منطقة ما، أو متوسط إنتاج الفدان لمحصول من المحاصيل الزراعية، أو متوسط دخل الفرد من الدخل القومي في بلد ومقارنة ذلك ببلدان أخرى.

ومن المعروف أن المتوسط الحسابي لمجتمع ظاهرة ما يعد قيمة ثابتة، بينما تعد قيمة المتوسط الحسابي لعينة قيمة متغيرة تختلف من عينة إلى أخرى للمجتمع الواحد. فمثلاً إذا سحبنا عدداً من العينات من مجتمع ما فإن المتوسط الحسابي غالباً ما يختلف من عينة إلى أخرى. ولو أن المتوسط الحسابي لمتوسطات هذه العينات يمكن اعتباره تقدير لمتوسط المجتمع. وفي معظم الحالات يعتبر المتوسط الحسابي للعينة قيمة غير متحيزة Unbiased لمتوسط المجتمع. أو بمعنى آخر أن المتوسط الحسابي لأي عينة قد يزيد أو يقل عن متوسط المجتمع الذي سحبت منه العينة إلا أن متوسط متوسطات العينات يطابق المتوسط العام للمجتمع. وسنعود لدراسة هذا الموضوع بالتفصيل فيما بعد.

المتوسط الحسابي للبيانات غير المبوبة:

يمكن تعريف المتوسط الحسابي لمجموعة من المفردات على أساس أنه القيمة التي لو أعطيت لكل مفردة من مفردات البيانات لكان مجموع هذه القيم الجديدة هو نفس مجموع المفردات الأصلية. كما يعرف المتوسط الحسابي (حسابياً) على أساس أنه القيمة الناتجة من جمع قيمك المفردات كلها مقسوماً على عدد المفردات. فمثلاً لو كان لدينا مجموعة (ن) من المفردات $s_1, s_2, s_3, \dots, s_n$ إلخ، فإن متوسطها الحسابي والذي يرمز له (س) يمكن حسابه كالآتي:

$$\bar{s} = \frac{\text{مجموع قيم المفردات}}{\text{عدد المفردات}} = \frac{s_1 + s_2 + s_3 + \dots + s_n}{n}$$

$$س = \frac{مجدس}{ن} \dots \dots \dots (٤ - ١)$$

والمعادلة الجبرية السابقة يمكن تطبيقها في كل الحالات التي تكون فيها البيانات ذات قيم مفردة. لذا فإنها تسمى بالأسلوب المباشر لحساب المتوسط الحسابي.

مثال تطبيقي:

لإيجاد المتوسط السنوي للأمطار في كل من مدينتي سيدني Sydney (استراليا) وستانلي مور Stanley Moor (انجلترا) خلال فترة ٢٠ سنة الموضحة بالجدول رقم (٤-١) تجرى الخطوات التالية.

١- تجمع كمية المطر (س) العشرين لكل مدينة على حدة.

٢- يقسم المجموع الكلي على عدد السنين (ن) لكل من المدينتين.

جدول رقم (٤-١)

كمية المطر السنوي (ستيمتر) في سيدني (استراليا)

وستانلي مور (انجلترا) في الفترة من ١٩٣٣ - ١٩٥٢

السنه	سيدنى	ستانلى مور	السنه	سيدنى	ستانلى مور
١٩٣٣	١٠٨	١٠٦	٤٤	٧٩	١٠٤
٣٤	١٦٥	١٣٨	٤٥	١٨٠	١٤٤
٣٥	٧٩	١٢٥	٤٦	٩٢	١٠٨
٣٦	٧٧	١٠٣	٤٧	١٠٥	١٥٢
٣٧	١٣٢	١٢٨	٤٨	٩٩	١١٩
٣٨	٩٩	١٣٢	٤٩	١٦٨	١٣٥
٣٩	٨٥	١١٨	٥٠	٢١٩	١٥٥
٤٠	١٠٠	١١٧	٥١	١٣٥	١٣٤
٤١	٦٨	١٢٠	١٩٥٢	١٥٠	١١٦
٤٢	١٢٣	١١٤			
٤٣	١٢٩	١٣٠	المجموع	٢٣٩٢	٢٤٩٨

وعلى ذلك يكون المتوسط الحسابي للمطر في كل مدينة هو:

$$\text{س (سيدنى)} = \frac{\text{مجم س}}{\text{ن}} = \frac{2392}{20} = 119,6 \text{ سنتيمتر}$$

$$\text{س (ستانلى مور)} = \frac{\text{مجم س}}{\text{ن}} = \frac{2498}{20} = 124,9 \text{ سنتيمتر}$$

وهذا يفسر جغرافياً على أنه على الرغم من اختلاف كمية المطر السنوية من مدينة لأخرى إلا أنه يوجد اتجاه عام وهو أن متوسط كمية المطر خلال فترة العشرين عاماً هي 119,6 سنتيمتر فى سيدنى و 124,9 سنتيمتر فى ستانلى مور. ويعد هذا المؤشر من خصائص مجتمع المطر فى كل مدينة من المدينتين.

وفى حالة وجود عدد كبير من المفردات لا ينصح باستخدام الأسلوب المباشر فى قياس المتوسط الحسابى حيث يطول وقت العمليات الحسابية والتى بالتالى تكون عرضه لاحتمالات الخطأ. ولكن ينصح باستخدام أسلوباً مختصراً لقياس المتوسط الحسابى يعتمد على اختصار الوقت وتبسيط العمليات الحسابية وتقليل درجة الأخطاء الممكنة، مع الحصول على نفس النتائج التى يمكن التوصل إليها باتباع الأسلوب المباشر. ويستعين الأسلوب المختصر بإحدى الطرق التالية لتحديد المتوسط الحسابى: طريقة العامل المشترك أو طريقة الوسط الفرضى. وفيما يلى خطوات إجراء كل منهما.

١ - طريقة العامل المشترك:

تعتمد هذه الطريقة على قسمة القيم الأصلية للمفردات على مقدار ثابت يطلق عليه اسم «العامل المشترك» وذلك لتبسيط العمليات الحسابية. وتقوم فكرة العامل المشترك على أساس أن المتوسط الحسابى لمجموعة من مفردات البيانات يمكن الحصول عليه لو أننا ضربنا ناتج قسمة مجموع المفردات المختصرة (أى بعد قسمتها على العامل المشترك) على عددها فى هذا العامل المشترك. فمثلاً إذا كانت لدينا المفردات التالية:

س_١ ، س_٢ ، س_٣ ، ، س_ن

وعلى فرض أن (ل) تمثل العامل المشترك بين هذه المفردات، فإن قيم المفردات المختصرة (المفردات الأصلية مقسومة على العامل المشترك) تصبح كالآتي:

$$\frac{س_١}{ل} ، \frac{س_٢}{ل} ، \frac{س_٣}{ل} ، ، \frac{س_ن}{ل}$$

وعلى ذلك فإننا نحصل على المتوسط الحسابي بواسطة الصيغة التالية:

$$\text{المتوسط الحسابي} = \frac{\text{مجموع قيم المفردات المختصرة}}{\text{عددها}} \times \text{العامل المشترك}$$

$$س = \text{مجم} \times \frac{س}{ل} \times (٤ - ٢)$$

وهناك شرط يجب مراعاته عند أخذ العامل المشترك لمجموعة من المفردات وهو أن يمثل العامل المشترك مقدار تسهل معه العمليات الحسابية ولا يسبب وجود كسور عشرية في ناتج قسمة المفردات الأصلية.

مثال تطبيقي:

إذا كانت لدينا بيانات مجموعة من عشرة أشخاص عمر كل منهم على النحو التالي:

١٠ ، ٣٠ ، ٤٠ ، ٥٠ ، ٢٠ ، ٣٠ ، ٤٠ ، ١٠ ، ٣٠ ، ٢٠

والمطلوب: إيجاد متوسط السن لهذه المجموعة باستخدام طريقة العامل المشترك

فإننا نتبع الخطوات الآتية:

- ١- نختار عاملاً مشتركاً هو القيمة (١٠) في توزيع أعمار الأشخاص العشرة.
- ٢- نقسم مفردات البيانات (قيم الأعمار) على هذا العامل المشترك فنحصل بذلك على المفردات المختصرة للبيانات.
- ٣- نجمع القيم المختصرة ونقسم على عدد المفردات ثم يضرب الناتج في العامل المشترك حتى نحصل على المتوسط الحسابي.

$$\frac{20}{10} + \frac{30}{10} + \frac{40}{10} + \frac{10}{10} + \frac{30}{10} + \frac{20}{10} = \text{المتوسط الحسابى س} \\ \frac{10}{10} + \frac{30}{10} + \frac{40}{10} + \frac{50}{10}$$

$$10 \times [10 \div (1 + 3 + 4 + 5 + 2 + 3 + 4 + 1 + 3 + 2)] =$$

$$28 \text{ سنة} = 10 \times 2,8 = 10 \times \left(\frac{28}{10} \right) = \text{س} \therefore$$

٢- طريقة الوسط الفرضى:

تستخدم هذه الطريقة لتسهيل واختصار العمليات الحسابية لإيجاد المتوسط الحسابى لعدد المفردات. وتقوم فكرة هذه الطريقة على أساس اختيار أى قيمة من بين قيم المفردات كوسط فرضى ثم نوجد إنحرافات قيم المفردات جميعها عن الوسط الفرضى المختار. وبعد ذلك نوجد متوسط الإنحرافات عن الوسط الفرضى ويضاف ذلك إلى الوسط الفرضى نفسه الذى يرمز له بالرمز (أ)، فنحصل على المتوسط الحسابى الحقيقى للمفردات، ويتم ذلك بالصيغة الإحصائية الآتية:

$$\text{المتوسط الحسابى} = \text{الوسط الفرضى} + \frac{\text{مجموع إنحرافات قيم المفردات عن وسطها الفرضى}}{\text{عدد المفردات}}$$

فعلى فرض أنه لدينا مجموع المفردات (س_١ + س_٢ ، س_٣ ، سن) وعلى فرض أن (أ) هى الوسط الفرضى لهذه المجموعة، فإن الإنحرافات عنه تأخذ الشكل التالى:

$$(س_١ - أ) ، (س_٢ - أ) ، (س_٣ - أ) ، (سن - أ)$$

$$\text{أى أن المتوسط الحسابى} =$$

$$\frac{(س_١ - أ) + (س_٢ - أ) + (س_٣ - أ) + + (سن - أ)}{ن} + أ$$

وعلى ذلك فإن :

$$\bar{A} = \frac{\text{مجم (س - أ)}}{n} + \bar{A}$$

وبإعطاء الرمز (ح) للتعبير عن انحرافات قيم المفردات عن وسطها الفرضي (س - أ) فإن

$$\bar{A} = \frac{\text{مجم ح}}{n} + \bar{A} \quad \dots\dots \quad (3-4)$$

وعند استخدام طريقة الوسط الفرضي كأسلوب احصائي مختصر لإيجاد المتوسط الحسابي للبيانات فإنه يجب علينا مراعاة مايلي:

أ- أن لا يكون الوسط الفرضي أصغر أو أكبر من قيم المفردات حتي لا يصبح مجموع الانحرافات دائماً موجباً أو سالباً بل يفضل أن تكون قيمة الوسط الفرضي هي القيمة التي تجعل مجموع الانحرافات أقل ما يمكن.

ب- يفضل أخذ الوسط الفرضي من أحد قيم المفردات الأكثر تردداً أو تكراراً أو تلك التي تتوسط عدد المفردات من حيث القيمة.

مثال تطبيقي:

إذا اعتبرنا أن الوسط الفرضي هو (٣٠) في توزيع أعمار الأشخاص العشرة السابق ذكره، فإن إيجاد قيمة المتوسط الحسابي لهذا التوزيع بطريقة الوسط الفرضي يعتمد على تنظيم البيانات حسب الخطوات السابقة ووضعها على شكل جدول، كما هي الحال في الجدول التالي:

جدول رقم (٤-٢)

حساب المتوسط الحسابي بطريقة الوسط الفرضي

رقم الشخص	السن (س)	(س - أ)	(ح)
١	٢٠	٣٠ - ٢٠	١٠ -
٢	٣٠	٣٠ - ٣٠	صفر
٣	١٠	٣٠ - ١٠	٢٠ -
٤	٤٠	٣٠ - ٤٠	١٠
٥	٣٠	٢٠ - ٣٠	صفر
٦	٢٠	٣٠ - ٢٠	١٠ -
٧	٥٠	٣٠ - ٥٠	٢٠
٨	٤٠	٣٠ - ٤٠	١٠
٩	٣٠	٣٠ - ٣٠	صفر
١٠	١٠	٣٠ - ١٠	٢٠ -
			٤٠ +
			٦٠ -
			٢٠ -

وعلى ذلك فإن

$$\text{متوسط عمر الأشخاص} = ٣٠ + \left(\frac{٢٠ -}{١٠} \right) = ٢٨ \text{ سنة}$$

ومن الملاحظ أننا حصلنا على نفس الجواب السابق ولكن باستخدام أسلوب حسابي مبسط.

المتوسط الحسابي للبيانات المبوبة

لا تختلف عملية حساب المتوسط الحسابي كثيراً إذا كانت قيم المفردات تحدث بتكرارات، أى من نوع البيانات المبوبة، عنها مع البيانات غير المبوبة. غير أن في حالة التوزيعات التكرارية البسيطة للبيانات يجب ضرب كل قيمة في تكرارها المناظر حتى نحصل على مجموع القيم ويقسم الناتج على المجموع الكلى للتكرارات. ولنفرض أن لدينا توزيعاً تكرارياً يشتمل على:

قيم المفردات: $s_1, s_2, s_3, \dots, s_n$ من المجموع

التكرار: $k_1, k_2, k_3, \dots, k_n$ مع ك

فيكون المتوسط الحسابي هو
$$\frac{\text{مجموع (القيمة} \times \text{التكرار)}}{\text{مجموع التكرارات}}$$

$$= \frac{s_1 k_1 + s_2 k_2 + s_3 k_3 + \dots + s_n k_n}{k_1 + k_2 + k_3 + \dots + k_n}$$

وعلى ذلك فإن:

$$\bar{s} = \frac{\text{مجموع } s \times k}{\text{مجموع } k} = \frac{(s-k)}{(s-k)}$$

ويطلق في بعض الأحيان على التكرارات اسم معاملات الترجيح أو أوزان، وهذه تعتمد على الدلالة أو الأهمية المرتبطة بقيم المفردات حتى أن المتوسط الحسابي الذى نحصل عليه بالمعادلة (٤-٤) يعرف باسم المتوسط الحسابي المرجح

(الموزون) Weighted Arithmetic Mean

مثال:

الجدول الآتى يبين توزيع حجم ٢٠ أسرة المطلوب إيجاد متوسط حجم هذه الأسر.

جدول رقم (٤-٣)
حساب المتوسط الحسابي المرجح

حجم الأسرة × التكرار س × ك	عدد الأسر (التكرار) ك	حجم الأسرة
٣	٣	١
٨	٤	٢
٦	٢	٣
١٦	٤	٤
١٠	٢	٥
١٢	٢	٦
١٤	٢	٧
٩	١	٩
٧٨	٢٠	المجموع ٣٧

وبتطبيق الصيغة الرياضية (٤-٤) فإن:

متوسط حجم الأسرة المرجح (س) = $\frac{٧٨}{٢٠} = ٣,٩$ (أي ٤ أشخاص تقريباً)
بينما يكون المتوسط الحسابي العادي هو:

$$س = \frac{٣٧}{٨} = ٤,٦٢٥ \text{ (أي ٥ أشخاص تقريباً)}$$

وهنا يظهر أن المتوسط الحسابي العادي (٤,٦٢٥) يعد متوسط لا يمثل مركز البيانات الحقيقي. بينما نستنتج أن المتوسط الحسابي المرجح (٣,٩) أكثر ملائمة في هذه الحالة لاختلاف الأهمية أو الدقة النسبية للبيانات ممثلة في التكرارات المناظرة لقيم المفردات (حجم الأسر).

ولتبسيط عملية الحساب يمكن أن نستخدم وسط فرضي ونوجد انحرافات قيم المفردات عن هذا الوسط (كما سبق شرحه)، ثم نضرب الانحرافات في التكرارات (الأوزان) المناظرة وأخيراً. يؤخذ متوسط مجموع الانحرافات ويضاف على الوسط الفرضي فنحصل بذلك على المتوسط الحسابي المرجح.

المتوسط الحسابي (س)

$$= 1 + \frac{(س_1 - أ_1) ك_1 + (س_2 - أ_2) ك_2 - (س_3 - أ_3) ك_3 + \dots + (س_n - أ_n) ك_n}{ك_1 + ك_2 + ك_3 + \dots + ك_n}$$

$$س = 1 - \left(\frac{\text{مجمد } (س - أ) ك}{\text{مجمد } ك} \right)$$

وبالتعويض عن (س - 1) بالرمز ح فإن

$$س = 1 - \left(\frac{\text{مجمد ح ك}}{\text{مجمد } ك} \right) \dots\dots\dots (4-5)$$

فإذا اعتبرنا أن الوسط الفرضي (أ) للمثال السابق، هو القيمة (4) فإنه يمكن الحصول على الجدول الآتي:

جدول رقم (٤-٤)

حساب المتوسط الحسابي المرجح بطريقة الوسط الفرضي

حجم الأسرة س	عدد الأسر ك	ح = س-١	ح ك
١	٣	٣-	٩-
٢	٤	٢-	٨-
٣	٢	١-	٢-
٤	٤	صفر	صفر
٥	٢	١	٢
٦	٢	٢	٤
٧	٢	٣	٦
٩	١	٥	٥
المجموع	٢٠		١٩- ١٧+

$$\text{أى المتوسط الحسابي} = \text{الوسط الفرضي} + \frac{\text{مجم ح ك}}{\text{مجم ك}}$$

$$= 4 + \left[\frac{2-}{20} \right]$$

$$= 4 - 0,1$$

$$= 3,9 \text{ وهى نفس النتيجة السابقة}$$

المتوسط الحسابي لجدول التوزيعات التكرارية:

سبق القول أن لاختصار عدد كبير من المفردات فإننا نحولها إلى فئات ونسجل عدد قيم المفردات (التكرارات) فى كل فئة فنحصل على جدول تكرارى.

ولاشك أن حساب المتوسط الحسابي من مثل هذا الجدول يكون أسهل وأسرع عن جمع عدد القيم كلها. ولما كان المتوسط الحسابي يعتمد في قياسه على مجموع قيم المفردات فإن وجود القيم في شكل فئات ينفي حقيقة القيم ويظهرها في شكل مقادير تنحصر بين حدى الفئة مما يجعل من الصعب تحديد مجموع القيم تحديداً دقيقاً. لذلك فإنه للحصول على حقيقة القيم نفترض أن كل عدد معين من التكرارات يحدث في منتصف مدى الفئة وأن القيم تتركز في مراكز فئاتها. وعلى ذلك فإن الخطوة الأولى في حساب المتوسط الحسابي لتوزيع تكرارى (حسب الفئات) هي إيجاد مراكز فئات التوزيع كمايلي:

$$\text{مركز الفئة (م)} = \frac{\text{الحد الأعلى للفئة} + \text{الحد الأدنى للفئة}}{2}$$

ويتحويل الفئات إلى مراكز الفئات (م) فإننا نعتبر الأخيرة هي القيم التي نريد إيجاد متوسطها الحسابي، علماً بأن لكل منها تكراراً معيناً (مجموع القيم يساوى مجموع التكرارات كلها) وبذلك تكون خطوات العمل لحساب المتوسط الحسابي كمايلي:

١- تجميع القيم في فئات مناسبة وتحديد عدد مرات حدوث كل قيمة (التكرارات).

٢- تحديد مراكز الفئات (م).

٣- اختيار وسط فرضي (أ) من بين مراكز الفئات. ويستحسن أن يكون مركز الفئة التي تقابل أكبر تكرار.

٤- حساب انحرافات (ح) مراكز الفئات عن الوسط الفرضي. أى ح = (م - أ).

٥- ضرب كل انحرافات في التكرار المناظر له (ح × ك) ثم إيجاد حاصل الضرب مجـ (ح ك).

٦- قسمة المجموع الناتج في الخطوة ٤ على مجموع التكرارات $\left(\frac{\text{مجـ ح ك}}{\text{مجـ ك}}\right)$.

٧- إضافة ناتج خارج القسمة إلى الوسط الفرضي (أ) فنحصل على المتوسط الحسابي المطلوب.

٨- يمكن أن نقسم الانحرافات (خطوة ٤) على طول الفئة (ل) إذا كان التوزيع منتظماً (متساوياً) في فئاته. فنحصل على $\frac{ك}{ل}$ ويكون المتوسط الحسابي في هذه الحالة هو:

$$س = ١ + \left[\frac{\text{مجموع } ك \times \text{مجموع } ل}{\text{مجموع } ك} \right] \times (٤-٦)$$

وتسمى الصيغة الرياضية السابقة بطريقة الترميز Coding Method عند حساب المتوسط الحسابي. وهذه الطريقة مختصرة جداً ويجب استخدامها دائماً للبيانات التكرارية (المجمعة) عندما تكون أطوال الفئات متساوية. فإذا كانت لدينا بيانات عن توزيع مساحات المزارع في منطقة ما (بالفدان) والبالغ عددها ٧٨ مزرعة مثلاً وأردنا حساب المتوسط المساحي لها فستكون خطوات الحساب على النحو الموضح بالجدول (٤-٥).

جدول رقم (٤-٥)
توزيع مساحات المزارع فى منطقة ما بالفدان

الانحراف × التكرار ل/ح × ك	الانحراف عن الوسط الفرضى ل/ك	مراكز الفئات م	التكرار ك عدد المزارع	الفئات مساحة المزارع
٦-	٦-	٥,٥	١	١٠ - ١
١٥-	٥-	١٥,٥	٣	٢٠ - ١١
١٦-	٤-	٢٥,٥	٤	٣٠ - ٢١
١٨-	٣-	٣٥,٥	٦	٤٠ - ٣١
١٨-	٢-	٤٥,٥	٩	٥٠ - ٤١
٩-	١-	٥٥,٥	٩	٦٠ - ٥١
صفر	صفر	٦٥,٥	١٠	٧٠ - ٦١
٩	١	٧٥,٥	٩	٨٠ - ٧١
١٦	٢	٨٥,٥	٨	٩٠ - ٨١
١٨	٣	٩٥,٥	٦	١٠٠ - ٩١
١٦	٤	١٠٥,٥	٤	١١٠ - ١٠١
١٥	٥	١١٥,٥	٣	١٢٠ - ١١١
١٨	٦	١٢٥,٥	٣	١٣٠ - ١٢١
١٤	٧	١٣٥,٥	٢	١٤٠ - ١٣١
٨	٨	١٤٥,٥	١	١٥٠ - ١٤١
٨٢ - ١١٤ +			٧٨	المجموع

وباتباع الخطوات السابقة لإيجاد المتوسط الحسابى نلاحظ الآتى:

الوسط الفرضى للتوزيع = ٦٥,٥

مجموع التكرارات = ٧٨

مجموع الانحرافات = ٣٢ +

وحيث أنه يمكن إيجاد المتوسط بتطبيق المعادلة (٤-٦)، وبالتعويض من حسابات الجدول فإن:

$$\begin{aligned} \bar{S} &= 65,5 + \left(10 \times \frac{32}{78} \right) \\ &= 65,5 + 4,1 \\ &= 69,8 \text{ فداناً} \end{aligned}$$

٢- المتوسط الهندسي Geometric Mean

تتميز العلوم الإجتماعية، ومن بينها الجغرافيا، بأن بعض ظواهرها لا تتغير في نسق منتظم أو بمقادير متساوية من فترة لأخرى، بل أنها تتغير بمعدلات مختلفة ومتباينة. ففي الديموجرافيا (علم السكان) تؤكد الدراسة على أن التغير السكاني لا يأخذ شكل المتوالية العددية، أى أن السكان لا يتزايدون أو يتناقصون بعدد متساوى من فترة لأخرى، بل أن التغير السكاني لا يتزايدون أو يتناقصون بعدد متساوى من فترة لأخرى، بل أن التغير السكاني يأخذ شكل المتوالية الهندسية، أى بنسبة تتناسب مع عدد السكان كل سنة وبناء على ذلك فإنه لا يمكن استخدام المتوسط الحسابي في قياس التغير السكاني باعتباره مجموع قيم المفردات على عددها، بل يجب استخدام الجذر النوني لحاصل ضرب أعداد السكان لفترات من السنين وهو ما يعرف باسم المتوسط الهندسي. ومن ثم فإن المتوسط الهندسي الذى يرمز له بالرمز (هـ)، لمجموعة من المفردات (ذات قيم موجبة) عددها (ن) يساوى الجذر النوني لحاصل ضرب قيم هذه المفردات. فعلى فرض أن لدينا عدد من المفردات على النحو التالى:

س_١، س_٢، س_٣،، س_ن فيكون متوسطها الهندسي هو :

$$\text{المتوسط الهندسي (هـ)} = \sqrt[n]{س_١ \times س_٢ \times س_٣ \times \times س_ن}$$

ولتسهيل إيجاد قيمة (هـ) نستعين باللوغاريتمات فيكون:

$$\text{لو هـ} = \frac{1}{n} (\text{لوس}_1 + \text{لوس}_2 + \text{لوس}_3 + \dots + \text{لوس}_n)$$

$$\text{لو هـ} = \frac{\text{مجلوس}}{n} \dots\dots\dots (7-4)$$

بمعنى أن لوغاريتم المتوسط الهندسى عبارة عن المتوسط الحسابى للوغاريتمات القيم المستخدمة فى حساب المتوسط الهندسى تتبع ماسبق شرحه - سابقاً - فى إيجاد المتوسط الحسابى مستخدمين لوغاريتمات قيم المفردات. ومن الناحية العملية فإن المتوسط الهندسى أقل تأثراً فى حسابه بالقيم المتطرفة عن المتوسط الحسابى. ومن ثم فإن المتوسط الهندسى لمجموعة من مفردات البيانات يكون أقل من المتوسط الحسابى لنفس المفردات مالم تكن قيم المفردات جميعها متساوية. فمثلاً إذا كانت لدينا قيم المفردات الآتية والتى تمثل أطوال بعض الطرق بالكيلو مترات وفى منطقة ما:

$$123, 120, 138, 160, 140, 172, 250$$

وأردنا حساب المتوسط الهندسى فإننا نوجد لوغاريتمات هذه القيم وهى على الترتيب:

$$2,0899, 2,1139, 2,1399, 2,2041, 2,1461, 2,2355, 2,3979$$

ويتطبيق المعادلة (7-4) فإن :

$$\text{لو هـ} = \frac{2,3979 + 2,2355 + 2,1461 + 2,2041 + 2,1399 + 2,1139 + 2,0899}{7}$$

$$= 2,1867$$

وبالكشف فى جداول الأعداد المقابلة اللوغاريتمات (الرقم 2,0867) نجد أن المتوسط الهندسى:

$$\text{هـ} = 153,81$$

أما إذا قمنا بحساب المتوسط الحسابى لنفس القيم بالمعادلة (7-4) نجد أن:

$$س = \frac{1113}{\sqrt{159}} \text{ (وهو كما ذكرنا أكبر من المتوسط الهندسي).}$$

ولحساب المتوسط الهندسي من جدول التوزيع التكرارى للبيانات فإننا نتبع نفس الأسلوب الذى استخدمناه لحساب المتوسط الحسابى بعد أن نوجد لوغاريتمات مراكز الفئات. ثم حساب المتوسط الهندسي كمايلي:

$$\text{لوه} = \frac{\text{مجم ك لوس}}{\text{مجم ك}} \dots\dots\dots (٤-٨)$$

مثال: الجدول التالى يمثل توزيع نسب أسعار ١٠٠ سلعة والمطلوب حساب متوسطها الهندسي.

جدول رقم (٤-٦)

حساب المتوسط الهندسي لنسب أسعار ١٠٠ سلعة

الفئات	التكرار	مراكز الفئات	لوم	ك × لوم
-١١٥	١٠	١٢٠	٢,٠٧٩٢	٢٠,٧٩٢
-١٢٥	٢٠	١٣٠	٢,١١٣٩	٤٢,٢٧٨
-١٣٥	٤٠	١٤٠	٢,١٤٦١	٨٥,٨٤٤
-١٤٥	٢٠	١٥٠	٢,١٧٦١	٤٣,٥٢٢
-١٥٥	١٠	١٦٠	٣,٢٠٤١	٣٣,٠٤٢
المجموع	١٠٠			٢١٤,٤٧٧

وعلى ذلك فإن:

$$\text{لوه} = \frac{\text{مجم (ك × لوس)}}{\text{مجم ك}} = \frac{٢١٤,٤٧٧}{١٠٠} = ٢,١٤٤٧٧$$

ويكون المتوسط الهندسى لهذا التوزيع $139,6$

ويعد المتوسط الهندسى من أنسب وأصلح مقاييس المتوسطات التى تصف الاتجاه العام لمجموعة من النسب أو المعدلات وبصفة خاصة معدلات النمو، ومعدلات المواليد كما فى المثال التالى.

مثال : فيما يلى معدل النمو السنوى للسكان لكل عشر سنوات فى مدينة ما والمطلوب حساب متوسط معدل النمو فى عدد السكان.

السنة	١٩٣٠	١٩٤٠	١٩٥٠	١٩٦٠	١٩٧٠	١٩٨٠
معدل النمو	٣,٥	٢,٥	٢,٠	١,٥	١,٠	٠,٥

$$\therefore \text{متوسط معدل النمو} = \sqrt[6]{٠,٥ \times ١,٠ \times ١,٥ \times ٢,٠ \times ٢,٥ \times ٣,٥}$$

$$\text{لو هـ} = \frac{1}{6} (\text{لو } ٣,٥ + \text{لو } ٢,٥ + \text{لو } ٢,٠ + \text{لو } ١,٥ + \text{لو } ١,٠ + \text{لو } ٠,٥})$$

$$\times \frac{1}{6} =$$

$$(١,٦٩٨٩ + ٠,٠٠ + ٠,١٧٦١ + ٠,٣٠١٠ + ٠,٣٨٧٩ + ٠,٥٤٤١)$$

$$٠,١٨٦٣ = ١,١١٨٠ \times \frac{1}{6} =$$

$$\text{هـ} = ١,٥٣٥٧$$

كذلك يصلح المتوسط الهندسى لقياس تقدير عدد السكان فى سنوات ما بين التعدادات. وهنا يكون التغير فى عدد السكان متناسباً مع عدد السكان نفسه.

مثال: كان سكان مصر فى تعداد سبتمبر ١٩٦٠ هو ٢٥٩٨٤٢٠١ نسمة وكان تعدادها فى مايو سنة ١٩٦٦ هو ٢٩٩٤٣٨١٠ نسمة. فما هو تعداد سكان مصر فى منتصف الفترة بين التعدادين أى فى ١٩٦٣. فإذا حسبنا التعداد فى منتصف الفترة عن طريق المتوسط الحسابى فإنه يكون :

$$\text{س} = \frac{٢٩٩٤٣٨١٠ + ٢٥٩٨٤٢٠١}{٢} = ٢٧٩٦٤٠٠٥ \text{ نسمة}$$

ولكن يعنى هذا أن الزيادة فى عدد السكان كل عام يكون بقدر متساوى وهذا لا يكون صحيحاً لأن معدل النمو فى هذه الحالة يتزايد مع تزايد السكان، ولذا فإن التقدير الصحيح يتم بإستخراج المتوسط الحسابى الهندسى. وحيث أن عدد القيم هو ٢ فإن :

$$\sqrt{(29943810 \times 20984201)} = \text{هـ}$$

$$\text{لو هـ} = \frac{1}{\frac{1}{29943810} + \frac{1}{20984201}}$$

$$\text{لو هـ} = \frac{1}{\frac{1}{7,41466} + \frac{1}{7,47626}}$$

$$\text{لو هـ} = 7,44546$$

$$\text{هـ} = 27893000 \text{ نسمة}$$

ويتضح من كل مما سبق أن المتوسط الهندسى على الرغم من مميزاته فى إستخراج المتوسطات للنسب والمعدلات إلا أنه أكثر صعوبة فى طريقة الحساب والفهم من المتوسط الحسابى.

٣- المتوسط التوافقى Harmonic Mean

يعرف المتوسط التوافقى لمجموعة من قيم المفردات بأنه مقلوب المتوسط لمقلوبات هذه القيم. فإذا فرضنا أنه كان لدينا القيم $s_1, s_2, s_3, \dots, s_n$ سن وعددها n فإن متوسطها التوافقى هو :

$$ق = \frac{n}{\frac{1}{s_1} + \frac{1}{s_2} + \dots + \frac{1}{s_n}} = \frac{n}{\text{مجم} \frac{1}{s}} \quad (٩-٤)$$

فالمتوسط التوافقى للأعداد ٢٠٠، ٤٠٠، ٦٠٠، ٨٠٠، ١٠٠٠ هو :

$$\frac{0}{0.1142} = \frac{0}{\frac{1}{1000} + \frac{1}{800} + \frac{1}{600} + \frac{1}{400} + \frac{1}{200}} = \bar{Q}$$

$$437.83 =$$

أما إذا حسبنا المتوسط الحسابي لهذه القيم فإنه يكون

$$\bar{Q} = \frac{2000}{0} = \frac{1000 + 800 + 600 + 400 + 200}{0} = \text{س}$$

ويعتبر المتوسط التوافقي أنسب مقياس المتوسط لتمثيل الأثمان ومعدلات السرعة وتعطى مثلاً يوضح تفسير استخدام المتوسط التوافقي بدلاً من المتوسط الحسابي على النحو التالي :

مثال: لنفرض أن معدل سرعة الحركة على أحد الطرق الملاحية (٤٠٠ كيلو متراً) يختلف من جزء إلى آخر على الطريق، فيمكن لأي باخرة أن تقطع ١٠٠ ك.م الأولى بسرعة ٢٠ كيلومتراً في الساعة، ثم ١٠٠ ك.م الثانية بسرعة ٣٠ كيلومتراً في الساعة، ثم ١٠٠ ك.م الثالثة بسرعة ٤٠ كيلومتراً في الساعة ثم ١٠٠ ك.م الأخيرة بسرعة ٥٠ كيلومتراً في الساعة. فإذا حسبنا المتوسط الحسابي لسرعة الباخرة فإنه يكون :

$$\bar{Q} = \frac{140}{4} = \frac{50 + 40 + 30 + 20}{4} = \text{س} = 35 \text{ كيلومتراً / ساعة}$$

وهذا غير صحيح، أو أن المتوسط الحسابي يكون في هذه الحالة مضللاً إلى حد كبير، وذلك لأن الباخرة قطعت المسافة الأولى في خمس ساعات والمسافة الثانية في ثلاث ساعات وثلاث والمسافة الثالثة في ساعتين ونصف والمسافة الرابعة في ساعتين. أي أنها استغرقت وقتاً وقدره ١٢ ساعة و ٥٠ دقيقة. وعلى ذلك يكون متوسط السرعة هو :

$$\text{متوسط السرعة} = \frac{400}{12.83} = 31.17 \text{ كيلو مترا/ ساعة}$$

والم متوسط السابق هو نفسه المتوسط التوافقي لسرعات الباخرة

$$\bar{Q} = \frac{4}{\frac{1}{0.} + \frac{1}{4.} + \frac{1}{3.} + \frac{1}{2.}} = \frac{4}{\frac{1}{0.} + \frac{1}{4.} + \frac{1}{3.} + \frac{1}{2.}}$$

$$= 31.17 \text{ كيلومترا/ ساعة}$$

وللمتوسط التوافقي لمجموعة من القيم خاصية هامة وهي أنه أقل من كل من المتوسط الحسابي والمتوسط الهندسي مالم تكن جميع القيم متساوية. كما يمكن حساب المتوسط التوافقي لجداول التوزيعات التكرارية، وفيها يضاف عموداً جديداً إلى جدول التكرار يوضع فيه مقلوب مراكز الفئات $(\frac{1}{M})$ ، ثم يضرب هذه المقلوبات في التكرارات المناظرة $(\frac{1}{M} \times K)$ ، ثم يوجد حاصل الجمع $[\frac{1}{M}]$ ، ثم يقسم عليه، مجموع التكرارات $\frac{\text{مجم ك}}{\text{مجم}}$.

مثال: الجدول الآتي يبين طريقة إيجاد المتوسط التوافقي لتوزيع تكراري لسرعات التيارات المائية في ١٠٠ رافد نهري لأحد أحواض التصريف المائي

جدول رقم (٤-٧)

طريقة حساب المتوسط التوافقي لسرعة التيارات المائية ١٠٠ رافد نهري

فئات السرعة ستيمتر، ثانية	عدد الأنهار ك	مراكز الفئات م	مقلوب مراكز الفئات $\frac{1}{M}$	$K \times \frac{1}{M}$
١٢,٥ -	٢٠	١٥	٠,٠٦٦٠	١,٣٣٣
١٧,٥ -	٥٠	٢٠	٠,٠٥٠٠	٢,٥٠٠
٢٢,٥ -	٢٠	٢٥	٠,٠٤٠٠	٠,٨٠٠
٢٧,٥ -	١٠	٣٠	٠,٠٣٣٣	٠,٣٣٣
المجموع	١٠٠			٤,٩٦٦

ويكون المتوسط التوافقي للسرعات هو:

$$Q = \frac{100}{4.966} = 20.14 \text{ سنتيمترًا ثانية}$$

وعلى الرغم من الصعوبات التي تواجه حساب هذا المقياس وكذلك صعوبة فهم الاحصائية للمتوسط التوافقي إلا أنه يفضل على باقى المتوسطات فى مجال الجغرافية، إذا ما كانت المفردات موضع الدراسة منسوبة إلى ثابت معين مثل السرعة فى الساعة (أو الدقيقة أو الثانية) أو بصفة عامة معدلات التغير المنسوبة إلى أساس ثابت.

ثانياً: الوسيط Mediau

الوسيط هو المقياس الثانى من مقياس النزعة المركزية. ويعرف الوسيط لمجموعة من المفردات أو مجموع من التكرارات بأنه القيمة التى تتوسط تلك المجموعة أو ذاك المجموع. أى أنه القيمة التى تقسم المجموعة إلى قسمين متساويين، القسم الأول منها يشمل عدد القيم الأصغر منها، والقسم الآخر يشمل عدد القيم الأكبر منها. ويمكن إيجاد الوسيط عن طريق ترتيب القيم تصاعدياً أو تنازلياً على هيئة منظومة Array للحصول على القيمة الوسطى لهذا الترتيب إذا كان عدد القيم فردياً أو متوسط القيمتين فى منتصف المنظومة إذا كان عدد القيم زوجياً.

فلو فرض أنه لدينا مجموعة المفردات التالية والتى تمثل أطوال سبعة أودية بالكيلو مترات ٨٣، ٦٨، ٦٦، ٦٥، ٧٣، ٦٩، ٧١. ثم نقوم بترتيب هذه القيم تصاعدياً (أو تنازلياً) فنحصل على:

$$٨٣ ، ٧٣ ، ٧١ ، ٦٩ ، ٦٨ ، ٦٦ ، ٦٥$$

ويكون الوادى الوسيط هو صاحب الطول الرابع فى الترتيب أى ٦٩ إذ أن هناك ثلاثة أطوال أكبر منه وثلاثة أطوال أصغر منه. ولإيجاد ترتيب الوسيط من مثل هذه المفردات ذات العدد الفردى فإنه يكون:

$$\text{ترتيب الوسيط} = \frac{\text{عدد القيم} + 1}{2} = \frac{1 + n}{2} \dots\dots\dots (4-10)$$

أما إذا كان عدد قيم المفردات (ن) زوجياً فإن قيمة الوسيط (ط) تقع ما بين قيمة المفردة التي ترتيبها $(\frac{n}{2})$ وقيمة المفردة التي ترتيبها $(\frac{n}{2} + 1)$. أو بمعنى آخر أن ترتيب الوسيط يقع بين القيمتين المتوسطين في الترتيب وأن قيمته يمكن الحصول عليها من حساب المتوسط الحسابي لهاتين القيمتين كما يبدو من المثال التالي لمعرفة القيمة الوسيطة لعمر ثمانية من الأشخاص (بالسنة):

٢٠ ، ٢١ ، ٢٦ ، ٢٧ ، ٢٩ ، ٣٠ ، ٣٢ ، ٣٣

$$\text{وهنا نجد أن ترتيب الوسيط} = \frac{1 + n}{2} = \frac{1 + 8}{2} = 4,5$$

أى أن قيمة الوسيط تقع في منتصف المسافة ما بين المفردة الرابعة والخامسة أو قيمة الوسيط $= \frac{1}{2} (\text{القيمة التي ترتيبها } \frac{n}{2} + \text{القيمة التي ترتيبها } (\frac{n}{2} + 1))$

$$= \frac{1}{2} (\text{القيمة الرابعة} + \text{القيمة الخامسة}) .$$

$$\therefore \text{قيمة الوسيط (ط)} = \frac{1}{2} (27 + 29) = 28 \text{ سنة}$$

إيجاد الوسيط للبيانات المبوبة:

لإيجاد الوسيط للبيانات المبوبة في شكل جداول توزيعات تكرارية للفئات نتبع الخطوات التالية:

- ١- ترتب التكرارات على شكل تكرارات متجمعة صاعدة أو هابطة.
- ٢- يحدد ترتيب الوسيط بين التكرارات وهو $= \frac{\text{مجموع التكرارات}}{2}$

$$= \frac{\text{مجم ك}}{2} \text{ بصرف النظر ما إذا كانت «ك» فردية أو زوجية.}$$

٣- يستخدم ترتيب الوسيط لتحديد الفئة التي يقع بها الوسيط (من عمود التكرار المتجمع بلا جدول) وتسمى الفئة الوسيطة. ثم تحدد الحدود الحقيقية للفئة الوسيطة وتكراراتها الأصلية.

٤- تحدد التكرارات لجميع الفئات التي تسبق الفئة الوسيطة بالاستعانة بالتوزيع التكرارى المتجمع الصاعد.

٥- يستخدم قانون استخراج الوسيط وهو:

الوسيط = الحد الأدنى الحقيقى للفئة الوسيطة +

$$\left[\frac{\text{ترتيب الوسيط} - \text{التكرار المتجمع الصاعد السابق للفئة الوسيطة}}{\text{التكرار الأصلي للفئة الوسيطة}} \right]$$

× طول الفئة]

$$= f_r + \left(\frac{\frac{1}{2}(n) - K_{\text{من}}}{K_r} \right) \times L \dots\dots\dots (3 - 11)$$

ولبيان طريقة حساب الوسيط للتوزيع التكرارى نستخدم بيانات الجدول التالى الذى يبين كميات الأمطار الساقطة بالمليمترات فى ٤٠ مرصداً جويّاً فى اقليم ما:

جدول رقم (٤-٨)
حساب الكمية الوسيطة للأمطار الساقطة
فى ٤٠ مرصداً بالمليمتر

الأمطار مليمترات	عدد المرصد التكرار	المتجمع الصاعد	
		أقل من الحد الأعلى للفئات	التكرار المتجمع الصاعد
١١٨ - ١٢٦	٣	أقل من ١٢٦	٣
١٢٧ - ١٣٥	٥	أقل من ١٣٥	٨
١٣٦ - ١٤٤	٩	أقل من ١٤٤	١٨
١٤٥ - ١٥٣	١٢	أقل من ١٥٣	٢٩ (فئة الوسيط)
١٥٤ - ١٦٢	٥	أقل من ١٦٢	٣٤
١٦٣ - ١٧١	٤	أقل من ١٧١	٣٨
١٧٢ - ١٨٠	٢	أقل من ١٨٠	٤٠
المجموع	٤٠		

وهناك طريقتان لإيجاد الوسيط مثل هذا التوزيع التكرارى هما:

(١) الطريقة الأولى، باستخدام الاستكمال نفترض أن كميات الأمطار فى الجدول التكرارى السابق (رقم ٤-٨) تتوزع توزيعاً مستمراً (متصلاً). فى هذه الحالة فإن الوسيط هو الكمية التى تقع نصف التكرارات الكلية أعلاها والنصف الآخر أسفلها وترتيبها هو نصف مجموع التكرارات $(\frac{40}{2} = 20)$. وحيث أن مجموع تكرارات الفئات الثلاث الأولى (التكرار المتجمع الصاعد للفئات أقل من ١٤٤) هو $3 + 5 + 9 = 17$. وحى نحصل على الترتيب المطلوب ٢٠ فإننا نريد ٣ تكرارات من التكرارات الموجودة فى الفئة التالية وهى الفئة التى يقع فيها الوسيط (١٢ تكرار). وبما أن الفئة الرابعة أو الفئة الوسيطة (١٤٥ - ١٥٣) هى

فى الحقيقة تقابل الكميات ١٤٤,٥ - ١٥٣,٥ فإن الوسيط يقع فى $\frac{3}{12}$ من المسافة بين ١٤٤,٥ و ١٥٣,٥ أى أن الوسيط هو:

$$\begin{aligned} \text{الوسيط} &= ١٤٤,٥ + \frac{3}{12} (١٥٣,٥ - ١٤٤,٥) \\ &= ١٤٤,٥ + \frac{3}{12} (٩) \\ &= ١٤٦,٨ \text{ ملليمتر} \end{aligned}$$

(٢) الطريقة الثانية، باستخدام المعادلة (٤-١١). بما أن التكرارات المتجمعة الصاعدة للفئات الأولى والفئات الأربع الأولى هى على الترتيب ٢٩, ١٧ فإن الوسيط يقع فى الفئة الرابعة التى هى بالتالى الفئة الوسيطة، وبذلك فإن:

$$\begin{aligned} \text{الحد الأدنى الحقيقى للفئة الوسيطة (ف)} &= ١٤٤,٥ \\ \text{المجموع الكلى للتكرارات (ن)} &= ٤٠ \end{aligned}$$

مجموع التكرارات المتجمعة الصاعدة للفئات السبعة للفئة الوسيطة (ك) = ١٧

$$\text{التكرار الأصلى للفئة الوسيطة (ك)} = ١٢$$

$$\text{طول الفئة (ل)} = ٩$$

وتكون قيمة الوسيط هى:

$$\text{الوسيط} = \text{ف} + \left(\frac{\frac{1}{2} - (\text{ن})}{\text{ك}} \right) \times \text{ل}$$

$$= ١٤٤,٥ + \left(\frac{\frac{1}{2} - (٤٠)}{١٢} \right) \times ٩$$

$$= ١٤٤,٥ + \left(\frac{١٧ - ٢٠}{١٢} \right) \times ٩$$

$$= ١٤٦,٨ \text{ ملليمتر}$$

وهى نفس النتيجة التى حصلنا عليها بطريقة الاستكمال.

تعيين الوسيط بيانياً:

يختلف الوسيط عن المتوسط الحسابى فى أنه يمكن إيجاد قيمته عن طريق الرسم البيانى للمدرج التكرارى أو لمنحنيات التكرار المتجمع الصاعد أو الهابط. فلو فرض أننا نريد تعيين الوسيط بيانياً فى مثال كميات الأمطار السابق فإننا نقوم بتمثيل التوزيعات التكرارية فى الجدول رقم (٤-١) وذلك برسم مدرجها التكرارى أو المنحنى التكرارى المتجمع النسبى (الصاعد أو الهابط)، ثم نعين ترتيب الوسيط وقيمه على أى منها كما يلى:

فى الشكل رقم (٤-١) الذى يوضح المدرج التكرارى المقابل لكميات الأمطار فى المثال السابق، يمكن تعيين ترتيب الوسيط منه على أساس أن الوسيط هو الأ-حدائى السينى للخط هـ ط الذى يقسم المدرج التكرارى إلى مساحتين متساويتين. وحيث أن المساحة تقابل التكرار فى المدرج التكرارى، فإن الخط هـ ط يقسم المساحة الكلية بحيث تكون التكرارات على يمينه والتكرارات على يساره مساوية لنصف التكرار الكلية أو ٢٠. كما أنه يقسم الفئة الوسيطة إلى قسمين متناسب مساحتهما عكسياً مع مجموع التكرارات السابقة واللاحقة لترتيب الوسيط. فمثلاً المساحة ط هـ د ب تناظر التكرار ٣ الذى يضاف إلى التكرارات السابقة للوسيط وعددها ١٧ لتصبح ٢٠، والمساحة أ ح هـ ط تناظر التكرار ٩ الذى يضاف إلى التكرارات اللاحقة للوسيط وعددها ١١ لتصبح ٢٠ أيضاً وبهذا فإن:

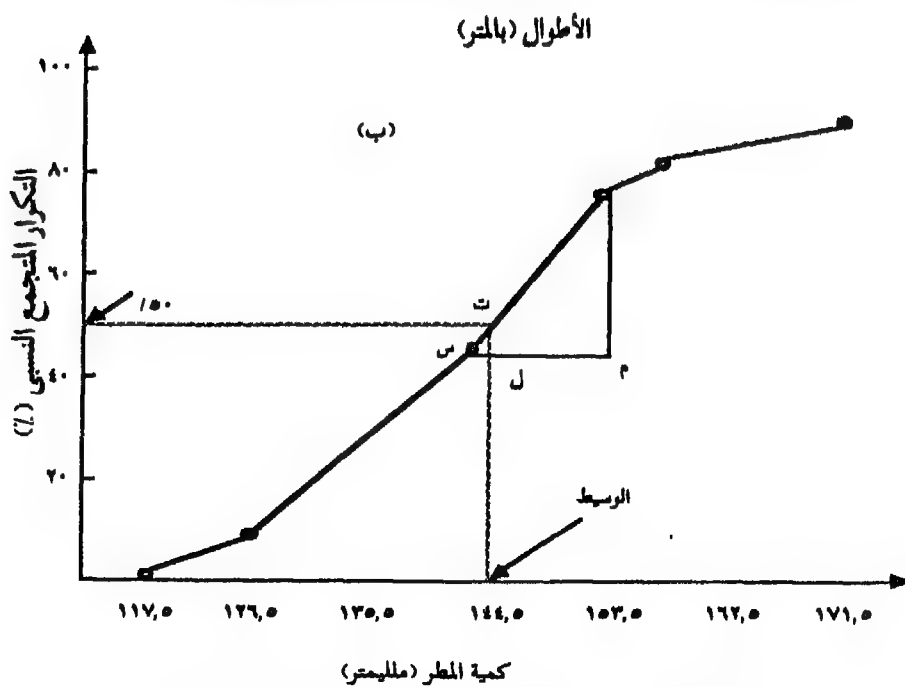
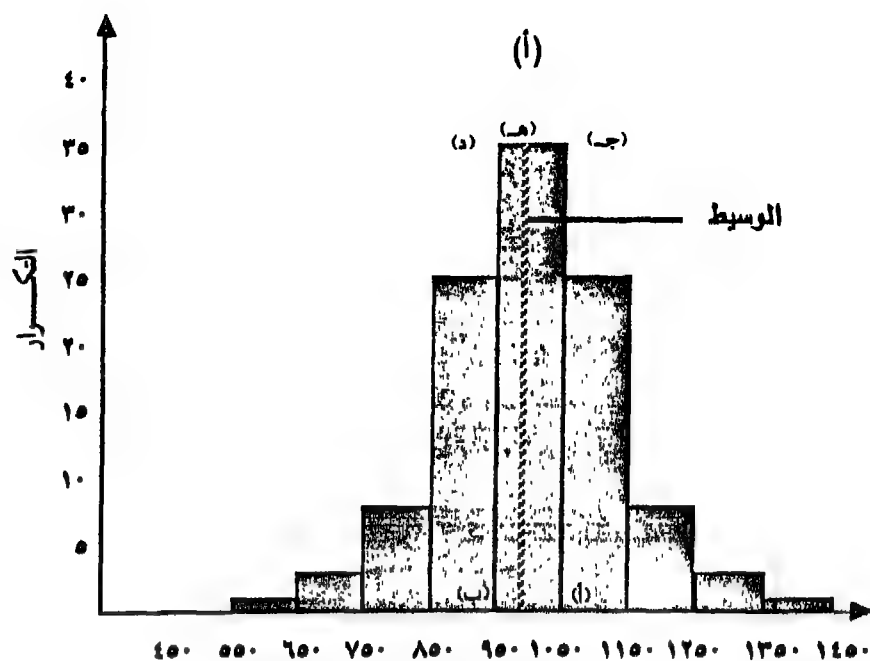
$$ط ب = \frac{٢}{١٧} (أ ب) = \frac{٢}{١٧} (٩) = ٢,٢٥$$

وتكون قيمة الوسيط (ط) = قيمة ب على الاحدائى السينى + قيمة ب ط

$$١٤٦,٨ = ٢,٢٥ + ١٤٤,٥ =$$

$$= ١٤٦,٨ \text{ ملليمتر (إلى أقرب نسبة من عشرة من الملليمتر)}$$

ويمكن قراءة هذه القيمة بشكل تقريبي مباشرة من الرسم.



شكل رقم (٤-١): تعيين الوسيط بيانياً من المدرج التكراري (أ) والمنحني المتجمع النسبي الصاعد

٣٠٥

أما إذا أردنا تعيين الوسيط من المحصى التكرارى المتجمع النسبى الصاعد المقابل لكميات الأمطار فى المثال السابق فإننا نقوم برسم محورين أحدهما رأسى يمثل التكرارات المتجمعة الصاعدة وآخر أفقى يمثل كميات الأمطار (الحدود الحقيقية للفئات) كما فى شكل رقم (٤ - ١ ب). وبما أن الوسيط هو القيمة التى يسبقها ويليها عدد متساو من التكرارات، فإننا نعين على المحور الرأسى القيمة التى تمثل نصف التكرارات الصاعدة (٥٠٪). ثم نمد خطاً أفقياً منها يتقاطع مع المنحنى التكرارى المتجمع فى نقطة (ت) التى يمثل احداثيتها السينى قيمة الوسيط. وللحصول على هذه القيمة فإننا نلاحظ من المثلثين المتماثلين ت ل س، ص م س أن:

$$\frac{ت ل}{ص م} = \frac{س ل}{م س}$$

وبما أن س م = طول الفئة = ٩

ت ل = (الاحداثى الصادى ت - الاحداثى الصادى ل)

ص ل = (الاحداثى الصادى ص - الاحداثى الصادى م)

فإن:

$$\frac{١}{٤} = \frac{٧,٥}{٣٠} = \frac{٤٢,٥ - ٥٠}{٤٢,٥ - ٧٢,٥} = \frac{س ل}{٩}$$

$$٢,٢٥ = \frac{٩}{٤} = س ل$$

وبهذا فإن الوسيط = ١٤٤,٥ + س ل

$$٢,٢٥ + ١٤٤,٥ =$$

$$١٤٦,٧٥ =$$

أو أن الوسيط = ١٤٦,٨ إلى أقرب عشر المليمتر. وهذه القيمة يمكن قراءتها بالتقريب من الرسم البياني.

شبهات الوسيط:

هناك مقاييس شبيهة بالوسيط تشترك معه في طريقه حسابها ولكنها ليست من المتوسطات مثل الربع Quartile والعشير Decile والمئين Centile. وسنكتفى بمعرفة طريقة حساب كل من الربع الأدنى والربع الأعلى نظراً لاستخدامها في حساب مقاييس التشتت فيما بعد.

الربع الأدنى والربع الأعلى:

يعرف الربع الأدنى Lower quartile بأنه تلك القيمة التي تقسم مجموعة البيانات إلى قسمين بحيث يسبقها ربع المفردات ويليهما ثلاثة أرباع المفردات، ويرمز له (Q₁). أما الربع الأعلى Upper quartile فيعرف بأنه القيمة التي تقسم البيانات إلى قسمين أيضاً بحيث تسبقها ثلاثة أرباع المفردات ويليهما ربع المفردات، ويرمز له (Q₃). ويمكن إيجاد ترتيب وقيمة كل من الربع الأدنى والأعلى للبيانات غير المبوبة والمبوبة باتباع نفس الطريقة التي استخدمناها لإيجاد الوسيط، كما يمكن إيجادهما بيانياً من رسم المنحنى التكرارى المتجمع الصاعد أو الهابط للبيانات. وستوجد الربعين الأدنى والأعلى لكميات الأمطار الساقطة على ٤٠ مرصد جوى (جدول رقم ٤-٨) كمايلي:

$$1 - \text{نوجد أولاً ترتيب الربع الأدنى وهو} = \frac{\text{مجموع التكرارات} \times 1}{4}$$

$$10 = \frac{1 \times 40}{4} =$$

٢- نحدد الفئة التي يقع الربع الأدنى من جدول رقم: (٤-٨) ونجدها الفئة (١٤٤ فأقل).

٣- نوجد قيمة الربع الأدنى باستخدام الصيغة الآتية:

$$\begin{aligned} & \text{الربيع الأدنى} = \text{الحد الأدنى لفئة الربيع الأدنى} + \\ & \left[\frac{\text{ترتيب الربيع الأدنى} - \text{التكرار المتجمع الصاعد السابق لفئة الربيع الأدنى}}{\text{التكرار الأصلي لفئة الربيع الأدنى}} \right] \\ & \times \text{طول الفئة} \end{aligned}$$

وبهذا فإن:

$$\text{الربيع الأدنى} = 136,5 + \left(\frac{8 - 10}{9} \right) \times 9$$

$$= 136,5 + 2 = 138,5 \text{ ملليمتر}$$

ومعنى هذا أن هناك عشرة مراصد تسقط عليها أمطار كميتها 138,5 ملليمتر أو أقل . وبنفس الطريقة يمكن إيجاد الربيع الأعلى كمايلي:

$$1 - \text{نوجد ترتيب الربيع الأعلى وهو} = \frac{\text{مجموع التكرارات} \times 3}{4}$$

$$= \frac{3 \times 40}{4} = 30$$

٢- نحدد فئة الربيع الأعلى من جدول التوزيع التكرارى (جدول رقم ٤-٨) فنجدها الفئة (١٥٤ فأقل)

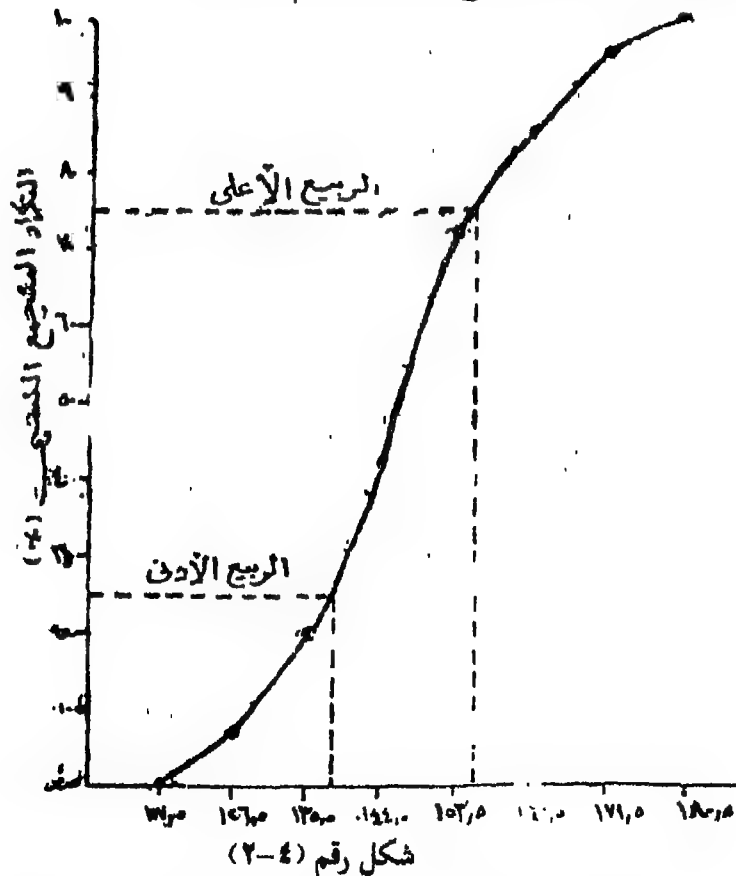
٣- نوجد قيمة الربيع الأعلى باستخدام نفس الصيغة السابقة لإيجاد الربيع الأدنى.

$$\text{الربيع الأعلى} = 154,5 + \left(\frac{29 - 30}{5} \right) \times 9$$

$$= 154,5 + 1,8$$

$$= 156,3 \text{ ملليمتر}$$

ولإيجاد الربيعين الأدنى والأعلى بيانياً نتبع نفس الخطوات التى انبعت لإيجاد الوسيط من منحنى التكرار المتجمع الصاعد أو الهابط فمن الشكل رقم (٢-٤) يمكن تحديد قيمة الربيع الأدنى بأن نعين ترتيبه على المحور الرأسى ونرسم خطاً أفقياً ليقابل المنحنى فى نقطة يمثل احداثيتها السينى القيمة المطلوبة وذلك عن طريق أسقاط عمود من هذه النقطة ليقابل المحور الأفقى عند قيمة ١٣٨,٥ ملليمتر تقريباً. وبالنسبة لإيجاد قيمة الربيع الأعلى فإننا نعين ترتيبه على المحور الرأسى ونتبع نفس الطريقة لتحديد قيمته على المحور الأفقى فنجد أن قيمته تساوى ١٥٦,٥ ملليمتر تقريباً كما هو واضح فى شكل رقم (٢-٤) :



تعيين الربيع الأدنى والأعلى بيانياً من المنحنى المتجمع النسبى الصاعد

ثالثاً: المنوال Mode

المنوال، أو ما يعرف أحياناً بالشائع، وهو أحد مقاييس النزعة المركزية التى تعطى تصوراً عاماً لتركز قيم مجموعة من المفردات حول قيمة متوسطة تمثل اتجاهها لهذه المجموعة كما تعتبر أحد خصائصها ومعالمها. ويمكن أن يعرف المنوال بأنه القيمة التى تحدث أكثر من غيرها فى مجموعة من البيانات أى هو القيمة الأكثر تكراراً أو شيوعاً فى التوزيع. فإذا كانت لدينا البيانات التالية للأحجام السكانية (بالألف نسمة) لعدد من المدن فى إقليم ما:

٧، ١٥، ٦، ٢٥، ٧، ٥، ٢٦، ٧، ١٥، ١٦

نجد أن القيمة (٧) تكررت أكثر من غيرها ولذلك فإن المنوال لهذه المجموعة يكون مساوياً للمدينة التى حجم سكانها ٧ آلاف نسمة. ويطلق على مثل هذا التوزيع الذى به قيمة واحدة هى الأكثر تكراراً إسم التوزيع أحادى المنوال Unimodal أما إذا كانت هناك قيمتان متساويتان فى تكرارهما فإن التوزيع يسمى بالتوزيع مزدوج (ثنائى) المنوال Bimodal كما هو واضح فى البيانات التالية للأحجام السكانية (بالألف نسمة) لعدد من المدن فى إقليم آخر:

٤، ٤، ٥، ٥، ٥، ٦، ٧، ٩، ٩، ٩، ١٠، ١١

يلاحظ أن كل من القيم ٩، ٥ هما القيمتان الأكثر تكراراً من غيرها من القيم الأخرى أى أن البيانات السابقة منوالين يتمثلان فى المدينتين اللتين حجم سكانها ٩، ٥ آلاف نسمة. وقد لا يوجد قيمة منوالية فى مجموعة البيانات عندما لا تتكرر قيمة ما أكثر من غيرها كما يلاحظ من بيانات الحجم السكانى (بالألف نسمة) لعدد من المدن فى إقليم ثالث.

٨، ٤، ١٣، ٩، ٧، ١٢، ٥، ١٠

إيجاد المنوال للبيانات المبوبة:

تختلف قيمة المنوال المحسوبة من التوزيع التكرارى عن قيمته المحسوبة من البيانات غير المبوبة. ويحسب المنوال من البيانات المبوبة فى جداول التوزيعات

التكرارية وذلك بتحديد الفئة التي تضم أكبر عدد من التكرارات وتعرف بالفئة المنوالية Modal Class ويمكن اعتبار مركز هذه الفئة منوالاً للتوزيع في حالة تساوى التكرارات المناظرة للفئتين قبل وبعد الفئة المنوالية.

وهناك عدة طرق لتحديد قيمة المنوال داخل الفئة المنوالية. وتعتمد الطريقة الأولى في حساب المنوال على معرفة تكرار كل من الفئتين المحيطين بالفئة المنوالية، وبالاقتباس من طريقة الرافعة فإن الفئة المنوالية تمثل الرافعة ويمثل تكرار الفئة قبل المنوالية القوة وتكرار الفئة بعد المنوالية المقاومة. وعلى هذا الأساس يتحدد موضع (قيمة) المنوال عند نقطة إرتكاز هذه الرافعة كما في الشكل رقم (٤-٣) اعتماداً على المثال التالي:

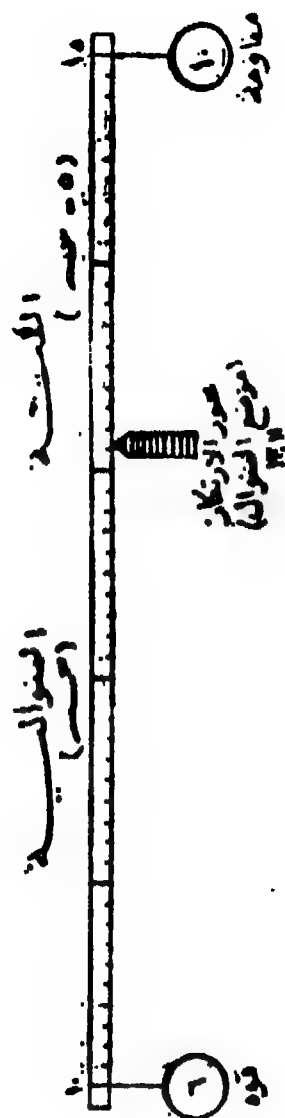
جدول رقم (٤-٩)

حساب المنوال لمساحة الأحواض

الزراعية المزروعة قطناً في أحد المحافظات «بالفدان»

مراكز الفئات	التكرار (عدد الأحواض)	الفئات (المساحة بالفدان)
٢,٥	٤	-٩
٧,٥	٦	-٥
١٢,٥ (الفئة المنوالية)	١٨	-١٠
١٧,٥	١٠	-١٥
٢٢,٥	٨	-٢٠
٢٧,٥	٢	-٢٥

فالأحواض الزراعية التي مساحتها من ١٠ لأقل من ١٥ فداناً هي الفئة المنوالية حيث أنها الأكثر تكراراً (١٨ تكراراً) والمنوال لهذا التوزيع هو ١٢,٥ حيث أنها القيمة المركزية للفئة ١٠ لأقل من ١٥. ولتحديد موضع (قيمة) المنوال داخل هذه الفئة يطبق قانون الرافعة.



شكل رقم (٣-٤)
تحديد موضع المنوال داخل الفئة المنوالية
«بطريقة الرافعة»

من الشكل السابق نرى أن المتوال يبعد مسافة قدرها s عن بداية الفئة (١٠) وبالتالي سيبعد مسافة قدرها (طول الفئة - s) أى (٥ - s) عن نهاية الفئة (١٥)، ومن قانون الرافعة نجد أن:

$$\text{القوة} \times \text{ذراعها} = \text{المقاومة} \times \text{ذراعها}$$

$$\text{التكرار السابق للفئة المتوالية} \times s = \text{التكرار اللاحق للفئة المتوالية} \times (٥ - s)$$

$$٦ \times s = ١٠ \times (٥ - s)$$

$$\therefore ٦s = ٥٠ - ١٠s$$

$$٦ = ٥٠$$

$$\therefore s = \frac{٥٠}{١٦} = ٣,١١$$

وعلى ذلك فإن:

$$\text{المتوال} = \text{الحد الأدنى للفئة المتوالية} + \text{قيمة محور الارتكاز (س)}$$

$$= ١٠ + ٣,١١ = ١٣,١١ \text{ فداناً}$$

أما الطريقة الثانية لحساب المتوال من جداول التوزيعات التكرارية فقد اقترحها بيرسون K. Pearson وتعرف بطريقة الفروق (بيرسون). وتقوم هذه الطريقة على أساس أن الذى يحدد موضع (قيمة) للمتوال داخل الفئة المتوالية هو الفرق بين تكرار الفئة المتوالية وتكرار الفئتين السبقة واللاحقة لها. وعلى ذلك يتحدد موضع المتوال بنسبة الفرق بين التكرارات على طرفي الفئة المتوالية وباستخدام العلاقة.

$$\text{المتوال} = \text{ف}_م + \left[\frac{ك - ك_١}{(ك - ك_١) + (ك - ك_٢)} \right] \times \text{ل} \dots (٤-١٢)$$

حيث أن:

$$\text{ف}_م = \text{الحد الأدنى للفئة المتوالية.}$$

- ك = تكرار الفئة المتوالية.
 ك_١ = تكرار الفئة السابقة لفئة المتوال.
 ك_٢ = تكرار الفئة اللاحقة لفئة المتوال.
 ل = طول الفئة.

ولإيجاد المتوال بهذه الطريقة من الجدول رقم (٤-٩) نجد أن :

الفرق بين تكرار الفئة المتوالية والفئة السابقة لها (ك - ك_١) = ١٨ - ٦ = ١٢
 الفرق بين تكرار الفئة المتوالية والفئة اللاحقة لها (ك - ك_٢) = ١٨ - ١٠ = ٨
 وبحسب المتوال على أساس أنه القيمة التي تقسم الفئة المتوالية (١٠ إلى أقل من ١٥) بنسبة ١٢ : ٨. فإذا كانت المسافة التي يبعد بها المتوال عن بداية الفئة هي س فإن المسافة التي يبعد بها عن نهاية الفئة هي (طول الفئة - س) أي (٥-س). وبهذا فإن:

س : (٥ - س) يجب أن تكون كنسبة ١٢ : ٨ أي أن :

$$\frac{١٢}{٨} = \frac{س}{(٥-س)}$$

$$\therefore ٨س = ١٢(٥-س)$$

$$٨س = ٦٠ - ١٢س$$

$$\therefore ٢٠س = ٦٠$$

$$\therefore س = ٣$$

والمتوال في هذه الحالة = ١٠ + ٣ = ١٣ فدائاً.

ويمكن الحصول على هذه النتيجة مباشرة باستخدام الصيغة (٤-١٢) :

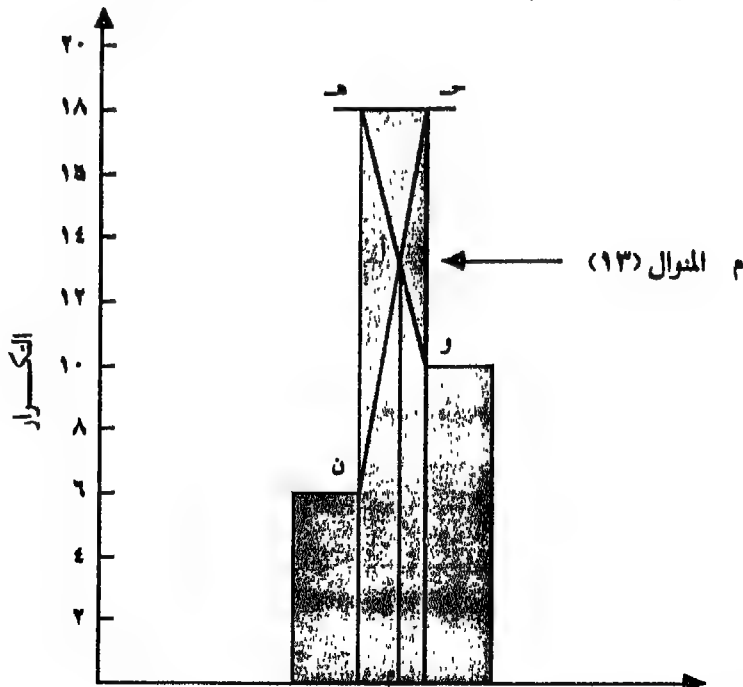
$$\text{المتوال} = ١٠ + \left(\frac{١٢}{٨+١٢} \right) \times ٥ = \frac{٦٠}{٢٠} + ١٠ = ١٣$$

$$١٣ = ٣ + ١٠ \text{ فدائاً}$$

إيجاد المنوال بيانياً:

المنوال مثل الوسيط يمكن الحصول عليه من الرسم البياني للمدرج التكرارى بتحديد الثلاثة مستطيلات التى يمثل المستطيل الأوسط منها تكرار الفئة المنوالية ويمثل كل من المستطيلين الآخرين تكرار الفئة السابقة والفئة اللاحقة لها على حدة (شكل رقم ٤-٤).

وبعين المنوال عن طريق توصيل طرفى المستطيل الذى يمثل تكرار الفئة المنوالية بطرفى المستطيلين اللذين يمثلان تكرارى الفئتين السابقة واللاحقة لها بمستقيمين، (ح د) و (هـ و)، أى بتوصيل الطرف الأيمن العلوى لمستطيل الفئة المنوالية بالطرف الأيمن العلوى للفئة السابقة لها (ح د)، والطرف الأيسر العلوى بالطرف الأيسر العلوى للفئة اللاحقة لها (هـ و)، والاحداثى السينى لنقطة التقاطع (أ) يساوى المنوال (م) الذى يمكن قراءة قيمته بشكل تقريبي مباشرة من الرسم كما هو موضح الشكل رقم (٤-٤).



شکل رقم ٤-٤ : طريقة تعيين المنوال بيانياً من المدرج التكرارى

العلاقة بين مقاييس النزعة المركزية:

سبق أن ذكرنا أن مقاييس النزعة المركزية الثلاثة: المتوسط الحسابي، الوسيط والمنوال هي أدوات القياس الكمي الرئيسية التي تعطى تصوراً عاماً لدرجة تركيز قيم المفردات حول قيمة متوسطة معينة وتعتبر هذه القيمة هي اتجاه عام لمجموعة مفردات البيانات وأحد معالمها وخصائصها. ومن التعريف السابق أمكن معرفة كل مقياس من هذه المقاييس، كما لوحظ أن هناك علاقة تربط بين كل منها. ولهذا فمن المفيد توضيح العلاقة هذه المتوسطات في ضوء الأمثلة الآتية:

فإذا أردنا مثلاً معرفة متوسط المساحة الزراعية المملوكة لعينة من عشرة زراع يمتلك كل منهم المساحة التالية (بالفدان) :

١، ٢، ٣، ٣، ٣، ٣، ٣، ٤، ٤، ٥

وتم تمثيل هذا التوزيع بيانياً في شكل مدرج تكرارى وممهداً عليه منحني توزيع تكرارى ليوافق هذا التوزيع (شكل رقم ٤-٥) فإننا نرى أن قيم المقاييس الثلاثة تتلاقى (تتطابق) في قيمة واحدة وهي الرقم (٣)، حيث أن:

$$\text{المتوسط الحسابي} = \bar{S} = \frac{\text{مجموع}}{n}$$

$$= \frac{30}{10} = 3 \text{ أفدنة}$$

$$\text{الوسيط} = \frac{2+3}{2} = 3 \text{ أفدنة}$$

المنوال = ٣ أفدنة (القيمة الأكثر تكراراً أو شيوعاً)

ويلاحظ من الشكل أيضاً أن منحني التوزيع التكرارى يعد متزناً تماماً على كلا جانبي قيم المتوسطات، ويعرف في هذه الحالة بمنحني التوزيع المتماثل الذى يعد، بالإضافة إلى تطابق قيم المقاييس الثلاثة عليه، سمة مميزة لما يعرف بالتوزيع التكرارى المعتدل (الطبيعى) Normal Frequency Distribution الذى تعتمد عليه

كثير من الأساليب الاحصائية على الرغم من ندرة حدوثه من الوجهة العملية. إذ كثيراً ما نجد أن قيم البيانات لا تشكل توزيعاً معتدلاً (متزناً) حول قيم المتوسط، بل تمثل توزيعاً غير معتدل Non- Normal Distribution مثل البيانات التالية التي تبين المساحات الزراعية (بالفدان) المملوكة لعينة من عشرة زراعي:

١، ١، ٢، ٢، ٢، ٣، ٤، ٤، ٥

فإننا نجد أن المتوسطات الثلاثة لهذه البيانات تختلف في قيمتها وذلك على النحو التالي:

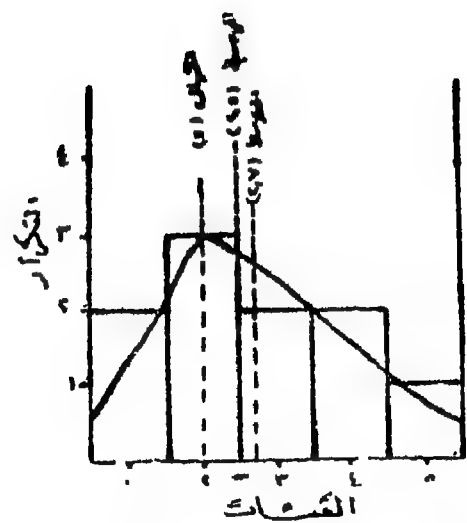
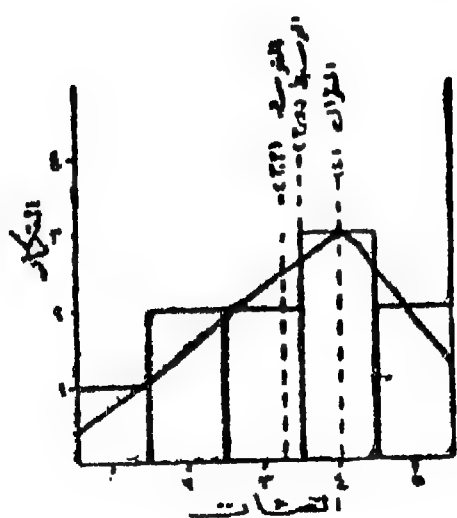
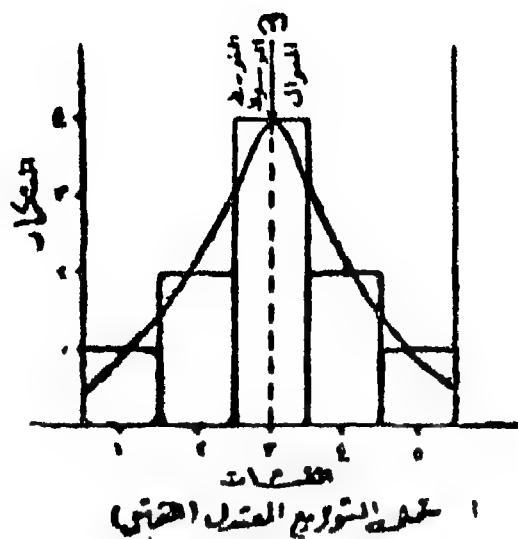
$$\text{المتوسط الحسابي} = \bar{X} = \frac{\text{مجموع}}{n} = \frac{27}{10} = 2.7 \text{ فداناً}$$

أما استخراج الوسيط والمتوال فيكون عن طريق جدول القيم أي إعادة ترتيبها أو تبويبها على النحو التالي:

القيمة	عدد مرات حدوثها (التكرار)
١	٢
٢	٣
٣	٢
٤	٢
٥	١

فالوسيط كما ذكرنا هي القيمة الوسطى. وبما أن عدد القيم يساوي ١٠، فإن الوسيط هو متوسط القيمتين الوسطيتين:

$$\text{الوسيط} = \frac{2 + 3}{2} = 2.5 \text{ فداناً}$$



٢. تخطيط التوزيع المتساوي

٣. تخطيط التوزيع المتساوي

شكل رقم (٤-٥)

مقاييس النزعة المركزية وعلاقتها بأنواع التوزيعات التكرارية

أما المتوال فهو القيمة التي تكررت أكثر من غيرها وهي القيمة (٢ فدان) التي تكرر حدوثها ثلاث مرات.

وقد تم توضيح القيم الثلاث السابقة على المدرج التكرارى والمنحنى التكرارى للتوزيع فى الشكل رقم (٤-٥ب) ويلاحظ من هذا الشكل أن المقاييس الثلاثة تبدو مختلفة عن بعضها ولا تتفق فى قيمة واحدة، كما رأينا فى المثال السابق، إذ يعد المتوسط الحسابى (٢,٧) أكبر قيمة بينما المتوال أصغرها (٢). كما يلاحظ أيضاً أن منحنى التوزيع غير متوازن (أى ملتو) - إذ أن له قمة تميل إلى الجانب الأيسر من المركز، وذيلاً يتجه صوب اليمين، ويسمى هذا بالالتواء الموجب - Positive Skewness. وتتميز التوزيعات مثل هذا النوع بأن النمط النسبى لتوزيع المتوسطات الثلاثة على منحنياتها يكون على نحو أن الوسيط يقع فى منتصف التوزيع بينما يقع المتوال على يساره والمتوسط الحسابى على يمينه كما يبدو من الشكل رقم (٤-٥ب).

كذلك تختلف المتوسطات الثلاثة فى قيمتها إذا كانت قيم البيانات تمثل توزيعاً غير معتدل وبصورة عكسية لبيانات المثال السابق. فمثلاً إذا كانت لدينا أيضاً بيانات بالمساحات الزراعية (بالفدان) المملوكة لعينة مكونة من عشرة من الزراع على النحو التالى:

١، ٢، ٣، ٣، ٤، ٤، ٤، ٥، ٥

فإن:

$$\text{المتوسط الحسابى} = \bar{س} = \frac{\text{مجم س}}{ن} = \frac{٣٣}{١٠} = ٣,٣ \text{ فدان}$$

أما بالنسبة لحساب لكل من الوسيط والمتوال فإن للقيم يعاد ترتيبها كمايلي:

القيمة	عدد مرات حدوثها (التكرار)
١	١
٢	٢
٣	٢
٤	٣
٥	٢

$$\text{الوسيط} = \frac{٣ + ٤}{٢} = ٢,٥ \text{ فداناً}$$

المنوال = ٤ فداناً (القيمة الأكثر تكراراً أو شيوعاً)

وهكذا تبدو أيضاً المتوسطات مختلفة عن بعضها البعض، فيعد المنوال أكبرها والمتوسط الحسابي أصغرهما، وهو عكس ما رأينا في التوزيع السابق. ومن الشكل رقم (٤-٥) الذي يوضح هذه القيم على مدرج تكرارى ممهداً عليه منحنى تكرارى للتوزيع، يبدو الأخير غير متوازن حيث يكون مائلاً نحو اليسار أن أى قمته تتجه إلى يمين المركز، وذيله صوب اليسار. وفي هذه الحالة يعرف بالتوزيع السالب الالتواء Negative Skewness. والتوزيعات الملتوية السالبة خاصية مميزة من حيث النمط النسبي لتوزيع المتوسطات الثلاثة على منحنياتها حيث نجد أن الوسيط يقع فى منتصف التوزيع والمتوسط الحسابي على يساره والمنوال على يمينه.

ومما تجدر الإشارة إليه هنا، أنه من الشكل رقم (٤-٥) يمكن استنتاج علاقة تقريبية عامة بين مقاييس النزعة المركزية الثلاثة للتوزيعات غير المتماثلة، وهذه العلاقة يمكن التعبير عنها حسابياً على النحو التالى:

أولاً: التوزيعات الملتوية الموجبة:

$$\text{المنوال} = \text{المتوسط الحسابي} - ٣ (\text{المتوسط الحسابي} - \text{الوسيط})$$

وفى تطبيق هذه العلاقة على المثال الموضح فى شكل رقم (٤-٥ ب) نجد أن:

$$\text{المنوال} = 2,7 - 3 (2,7 - 2,5)$$

$$= 2,7 - (3 \times 0,2)$$

$$= 2,7 - 0,6$$

$$= 2,1 \text{ فداناً}$$

وبقد سبق أن عرفنا أن القيمة للمنوالية من التوزيع هي (2)، والفرق بين القيمتين بسيط جداً، كما أن هذا يعنى أن الوسيط يقع إلى الخلف من المتوسط الحسابي بمقدار الثلث تقريباً وأمام المنوال بمقدار الثلثين تقريباً كما هو واضح في الشكل رقم (٤-٦).



شكل رقم (٤-٦)

العلاقة بين النزعة المركزية الثلاثة في التوزيعات التكرارية الموجبة الالتواء

جدول رقم (٤-١٠)

كمية الأمطار السنوية (بالبوصة) في مرصد

Bidston, Birkenhead - إنجلترا في الفترة ١٩٠١ - ١٩٣٠

السنة	كمية الأمطار بالبوصة	السنة	كمية الأمطار بالبوصة	السنة	كمية الأمطار بالبوصة
١٩٠١	٢٥,٩١	١٩١١	٢٥,٢٧	١٩٢١	٢٢,٤٧
٢	٢٥,٥٧	١٢	٣٠,١٧	٢٢	٢٥,٩٧
٣	٣٤,٤٢	١٣	٢٥,٧٨	٢٣	٣٠,٩٢
٤	٢٥,٨١	١٤	٢٦,٠٢	٢٤	٣٢,٨٧
٥	٢٤,٠١	١٥	٢٦,٨٣	٢٥	٢٨,٠٠
٦	٢٨,٠٨	١٦	٢٤,٨٧	٢٦	٢٨,٩٥
٧	٢٦,٥٧	١٧	٣٠,٥٩	٢٧	٣٤,٨١
٨	٢٨,٩٠	١٨	٣١,٩٣	٢٨	٢٩,١١
٩	٢٨,٣٥	١٩	٢٩,١٢	٢٩	٢٥,١٥
١٠	٢٨,٥٩	٢٠	٣٣,٣٤	٣٠	٣٦,٥٠

يبدو من بيانات الجدول السابق أن كمية الأمطار السنوية تتراوح من ٢,٢٤ بوصة و ٣٦,٥٠ بوصة، كما تمثل بيانات هذا الجدول متغيراً مستمراً (متصلاً) Continuous Variate. وبحساب المتوسط الحسابي من القيم في الجدول وجد أن:

$$\bar{x} = \frac{\sum x}{n} = \frac{٨٥٣,٦٣}{٣٠} = ٢٨,٤٥ \text{ بوصة}$$

ولحساب كل من الوسيط والمتوال فقد أعيد ترتيب البيانات في الجدول السابق في منظومة Array أى ترتيبها ترتيباً تصاعدياً حسب مقدارها، كما تم تبويبها أيضاً على شكل توزيع تكرارى بعد أن تم تجميع القيم في فئات مناسبة وعدد مرات

حدوث كل قيمة (التكرارات) أمام كل فئة. ولما كانت البيانات تمثل متغيراً متصلاً فإن كل القيم تشتمل على كسور عشرية، ومن ثم فلا بد أن تصمم حدود الفئات لتعطي متغيراً مستمراً فلا تكون مثلاً (٢١ - ٢٢)، (٢٣ - ٢٤) .. الخ، بل يجب أن تكون (٢١ إلى ٢٢،٩٩)، (٢٣ - ٢٤،٩٩) وهكذا. ويبين الجدول رقم (١١-٤) والشكل رقم (٤-٧) هذه البيانات.

جدول رقم (١١-٤)

كمية الأمطار السنوية (بال بوصة) في مرصد

Bidston, Birkenhead - إنجلترا في الفترة ١٩٠١ - ١٩٣٠

مرتبة بحسب مقدار القيمة

عدد مرات الحدوث التكرار	تحويل القيم إلى فئات	كمية الأمطار مرتبة حسب قيمتها
١	٢١ - ٢٢،٩٩	٢٢،٤٧
٢	٢٣ - ٢٤،٩٩	٢٤،٠١
١٠	٢٥ - ٢٦،٩٩	٢٤،٨٧
٦	٢٧ - ٢٨،٩٩	٢٥،١٥
٥	٢٩ - ٣٠،٩٩	٢٥،١٨
٢	٣١ - ٣٢،٩٩	٢٥،١٩
٣	٢٣ - ٣٤،٩٩	٢٥،٢٧
١	٣٥ - ٣٦،٩٩	٢٥،٥٧
الفئة المتوالية ٢٥ - ٢٦،٩٩		٢٥،٨٧
		٢٥،٩٧
		٢٦،٠٢
		٢٦،٥٧
		٢٦،٨٣
		٢٨،٠٠

ثانياً: التوزيعات المتتوية السالبة:

يمكن أن يعبر عن العلاقة التقريبية العامة التي تربط مقاييس النزعة المركزية الثلاثة (المتوسط الحسابي، الوسيط، المنوال) للتوزيعات السالبة الالتواء بصورة عكسية لنفس العلاقة مع التوزيعات المتتوية الموجبة وذلك على النحو التالي:

$$\text{المنوال} = \text{المتوسط الحسابي} + 3 (\text{الوسيط} - \text{المتوسط الحسابي})$$

ويتطبيق هذه المعادلة على بيانات الشكل رقم (٤-٥ ج) نجد أن:

$$\text{المنوال} = 3,3 + 3 (3,5 - 3,3)$$

$$= (0,2 \times 3) + 3,3 =$$

$$= 0,6 + 3,3 =$$

$$= 3,9 \text{ فدانا}$$

وسبق القول بأن القيمة المتتوية من للتوزيع هي القيمة (٤). والفرق بين القيمتين كما هو واضح بسيط جداً.

وهناك أيضاً علاقة تقريبية عامة بين المتوسطات الثلاثة يمكن بواسطتها حساب المتوسط الحسابي للتوزيعات غير المتماثلة (الموجبة أو السالبة). وهذه العلاقة يعبر عنها كمايلي:

$$\text{المتوسط الحسابي} = \frac{3 (\text{الوسيط}) - (\text{المنوال})}{2}$$

ويتطبيق هذه العلاقة على بيانات كل من الشكلين رقم (٤-٥ ب)، (٤-٥ ج) نجد أن:

$$١ - \text{المتوسط الحسابي} = \frac{(2) - (2,5 \times 3)}{2}$$

$$= \frac{0,5}{2} = 2,75 \text{ فدانا}$$

$$٢- \frac{(٤) - (٣,٥ \times ٣)}{٢} = \text{المتوسط الحسابى}$$

$$= \frac{٦,٥}{٢} = ٣,٢٥ \text{ فداناً -}$$

وكما نرى فإن القيمتين ٢,٧٥ و ٣,٢٥ قريبتان جداً من مثيلتهما فى التوزيع ٣,٣، ٢، ٧ على الترتيب مما يدل على دقة هذه العلاقة بين المتوسطات الثلاثة.

أمثلة تطبيقية:

بعد ذلك العرض العام الذى أوضحنا فيه العلاقة بين مقياس النزعة المركزية (المتوسط الحسابى، الوسيط والمتوال)، نحاول الآن إبراز هذه العلاقة بصورة حية وذلك بأخذ بيانات نوعية فى مجال الدراسات الجغرافية (الطبيعية والبشرية) وتطبيق الطرق السبق ذكرها لحساب هذه المقاييس الثلاثة عليها. فالجدول رقم (٤-١٠) يبين كمية الأمطار السنوية (بالبوصة) التى سجلها مرصد Bidston, Birkenhead - إنجلترا وذلك لمدة ثلاثين عاماً من ١٩٠١ إلى ١٩٣٠. والمطلوب تمثيل البيانات فى الجدول بيانياً، وإيجاد المتوسطات الثلاثة حسابياً وتوضيح ماينها من علاقة.

٢٨,٠٨
الوسيط (٢٨,٢٧)
٢٨,٤٥
٢٨,٥٩
٢٨,٩٠
٢٨,٩٥
٢٩,١١
٢٩,١٢
٣٠,١٧
٣٠,٥٩
٣٠,٩٢
٣١,٩٣
٣٢,٨٧
٣٣,٣٤
٣٣,٤٢
٣٤,٨١
٣٦,٥٠
المجموع = ٨٥٣,٦٣

ولما كان عدد قيم المفردات فى الجدول السابق ٣٠ قيمة فإن الوسيط سيتحدد فيما بين القيمتين للخامسة عشر والسادسة عشر، أى فى منتصف المسافة بين ٢٨,٨ بوصة و ٢٨,٤٥ بوصة أى أن:

$$\frac{28,45 - 28,08}{2} = \text{الوسيط}$$

وللمقارنة يمكن حساب الوسيط من التوزيع التكرارى للبيانات باستخدام الصيغة (١١-٤) فنجد أن:

$$\text{الوسيط} = 27,00 +$$

$$= 27,00 + (2 \times 0,33)$$

$$= 27,00 + 0,66$$

$$= 27,66$$

وكما نرى تختلف القيمة الأخيرة عن القيمة الدقيقة السابقة للوسيط. ويرجع هذا الخطأ - الذى يعتبر أكبر من الخطأ المعتاد فى مثل هذه الحالة - إلى وقوع كل التكرارات الأصلية الستة فى الفئة الوسيطة فى النصف الأعلى لهذه الفئة.

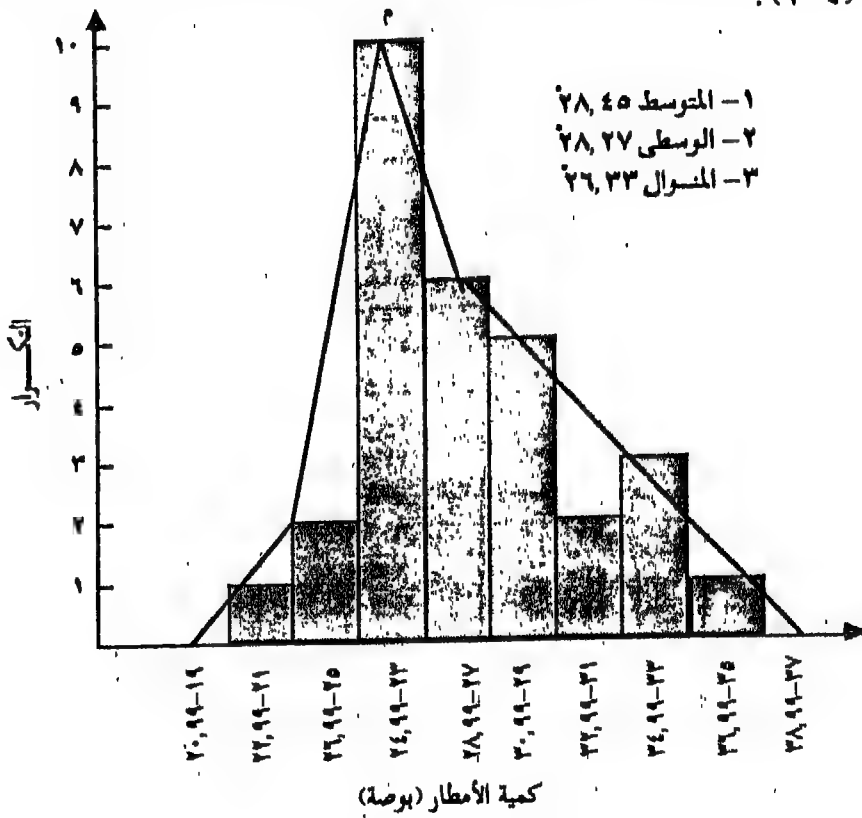
ومن الجدول رقم (١١-٤) والشكل رقم (٤-٧) يتضح أن الفئة ٢٥ - ٢٦,٩٩ هى الفئة المتوالية التى يوجد بها أكبر عدد للتكرارات (١٠). ويمكن بالتالى بعد تحديد الفئة التى يقع فيها المتوال حساب قيمته باستخدام العلاقة (١٢-٤) على النحو التالى:

$$\text{المتوال} = 25,00 + \left[\frac{2 - 10}{(7 - 10) + (2 - 10)} \right] \times 2$$

$$= 25,00 + 2 \times \frac{8}{12}$$

$$= 25,00 + 1,33 = 26,33 \text{ بوصة}$$

وهكذا تبدو المتوسطات الثلاثة مختلفة عن بعضها البعض، إذ نجد أن المتوسط الحسابي هو أكبر قيمة بينما المنوال أقلها. وقد تم تمثيل هذه القيم على المدرج التكراري والمنحني التكراري للتوزيع في الشكل رقم (٧-٤) ومنه يبدو أن المنحني ملتوئاً موجباً بسيطاً. ويستدل على ذلك، كما سبق القول من الصفة المميزة لكل التوزيعات الملتوية الموجبة وهي أن النمط النسبي للمتوسطات الثلاثة (أو توزيعها) على المنحني التكراري يأخذ شكلاً يكون فيه الوسيط في منتصف التوزيع والمنوال على اليسار والمتوسط الحسابي على اليمين ويظهر هذا بوضوح في الشكل رقم (٧-٤).



شكل رقم (٧-٤)

المدرج التكراري ومنحني التوزيع التكراري لكمية

الأمطار السنوية في مرصد ييستون - إنجلترا في الفترة ١٩٠١ - ١٩٣٠

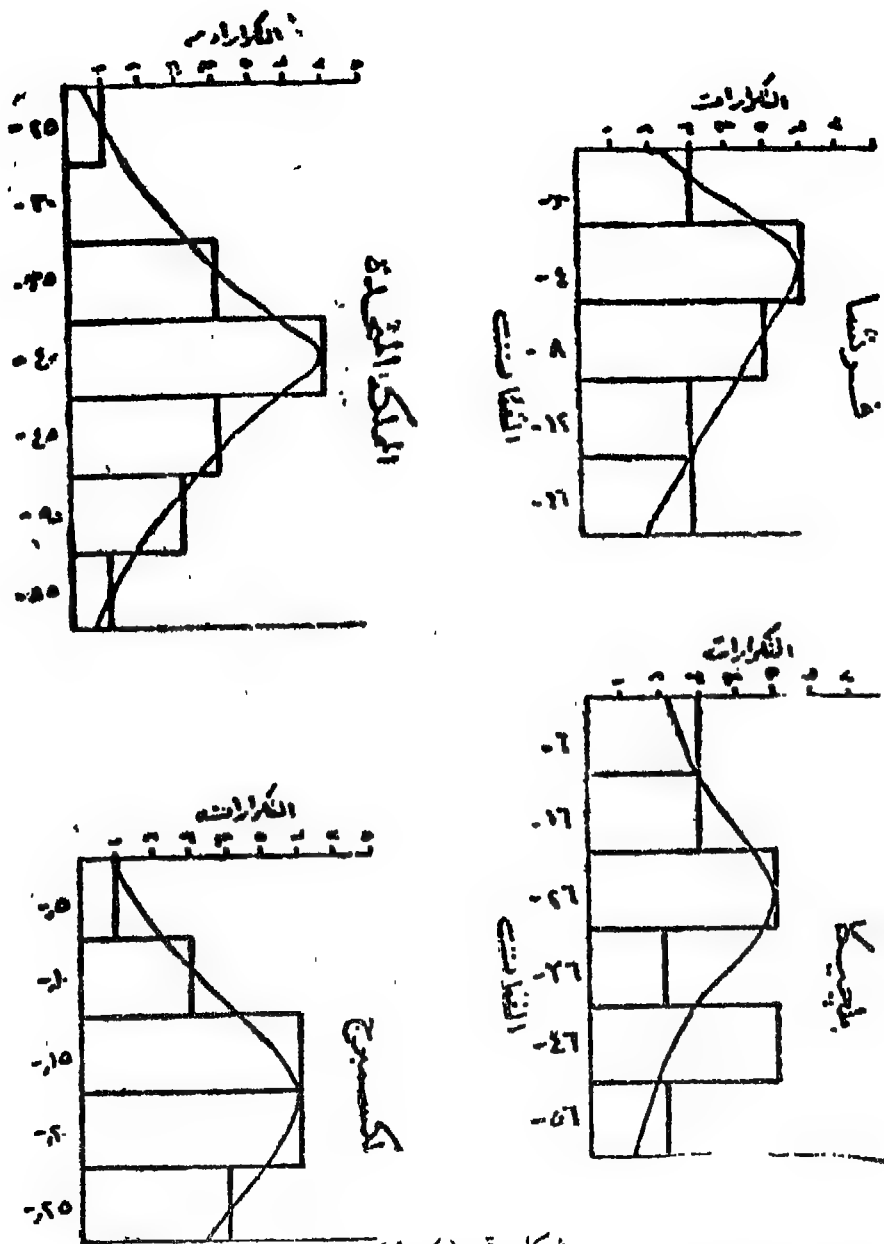
وتبدو الاختلافات فى التوزيعات التكرارية والعلاقة بين مقاييس النزعة المركزية المتوسطات الثلاثة فى بيانات تتصل بمجال الدراسة فى الجغرافية الاقتصادية، كما يظهر من بيانات الإنتاج السنوى لخام الحديد على مدى عشرين عاماً ١٩٣٨ - ١٩٥٧ فى أربع دول من غرب أوروبا هى: بلجيكا، فرنسا، لكسمبرج والمملكة المتحدة فى الجدول التالى:

جدول رقم (٤-١٢)

إنتاج خام الحديد بآلاف الأطنان فى

بعض دول غرب أوروبا فى الفترة من ١٩٣٨ - ١٩٥٧

السنة	بلجيكا	فرنسا	لكسمبرج	المملكة المتحدة
١٩٣٨	٦٥	١٠٢٠٣	١٥٠٦	٣٦١٥
٣٩	٦٠	١٠١٦١	١٦٣٩	٤٤١٧
٤٠	٢٩	٤١١٣	١٣٦٨	٥٤٤٩
٤١	٤٧	٣٤٦٧	١٩١٢	٥٥٢٨
٤٢	٤١	٤١٤٤	١٤٣١	٥٤٤٩
٤٣	٤٦	٥٣٥٠	١٤٧١	٥٤١١
٤٤	١٦	٢٨٦٢	٨١٦	٤٣٩٠
٤٥	١١	٢٣٤٢	٣٩٤	٤١٦٢
٤٦	١٤	٥٠٢١	٦٥٠	٣٥٧٤
٤٧	٢١	٦٠٩٩	٥٩٢	٢٩٨٤
٤٨	٣٤	٧٥٥٥	١٠٢٠	٣٩٩٠
٤٩	١٥	١٠٢٠٠	١٢٤١	٤٠٨٦
٥٠	١٦	٩٧٥٠	١١٥٤	٣٨١٢
٥١	٢٨	١١٤٤٠	١٦٨٨	٤٥٠٤
٥١	٤٧	١٣٢٣٠	٢١٧٤	٤٦١٨
٥٣	٣٥	١٣٧٩٠	٢١٥١	٤٥٠٠
٥٤	٢٩	١٤٢٤٠	١٧٦٦	٤٣٦٩
٥٥	٣٧	١٦٣٤٠	١٩٣٣	٤٤٣٧
٥٦	٥٠	١٧١٢٠	٢٠٣٤	٤٤٥٧
١٩٥٧	٤٨	١٨٧٧٠	٢٠٣٦	٤٦٣٧
المتوسط	٣٤,٤٥	٩٣١٠,٢	١٤٤٨,٨	٤٤١٨,٩٥
الوسيط	٣٤,٥	٩٩٥٥,٥	١٤٨٨,٥	٤٤٢٧,٠٠



المدرجات والمنحنيات التكرارية للإنتاج السنوي لخام الحديد في كل من بلجيكا، فرنسا، لكسمبرج والمملكة المتحدة في الفترة من ١٩٣٨ - ١٩٥٧

وكما هو واضح من الجدول السابق فقد تم حساب كل من المتوسط الحسابي والوسيط لإنتاج كل دولة بنفس الطرق السابق ذكرها مع بيانات الجدول رقم (٤-١١)، ويبدو في الحالات الأربع أن الوسيط أكبر من المتوسط الحسابي - مما يترتب عليه ميلاً بسيطاً نحو التواء سالب - بالرغم من أن التوزيع التكراري لبيانات فرنسا لا يؤكد ذلك. ويبين الشكل رقم (٤-٨) بيانات الدول الأربع. ويبدو من هذا الشكل أنه من الصعب تعيين منوال واضح من بيانات كل من بلجيكا ولكسمبرج بصفة خاصة، لوجود أكثر من فئة منوالية واحدة.

مزايا ومثالب مقاييس النزعة المركزية:

بعد ذلك العرض التفصيلي لطرق حساب كل مقياس من مقاييس النزعة المركزية الثلاثة وبعد معرفة العلاقة بينها، فإنه يمكننا الآن معرفة دلالة الفروق بين هذه المقاييس في الوصف الإحصائي واستنباط مزاياها وعيوبها كأدوات للقياس في التحليل الكمي للبيانات الجغرافية. فالمتوسط الحسابي كأحد مقاييس النزعة المركزية يكتسب أهمية خاصة بين الأنواع الثلاثة للمتوسطات باعتباره مقياساً دقيقاً يقوم على أسس رياضية سليمة، كما أن له مميزات وخصائص تسمح باستخدامه في إيجاد مقاييس أخرى نحتاج إليها في التحليل الإحصائي للبيانات. ومن بين هذه الخصائص أن مجموع انحرافات قيم المفردات عن المتوسط الحسابي، دون غيره من المقاييس الآخرين (الوسيط والمنوال)، يساوى صفراً، ويمكن إثبات هذه الخاصية جبرياً في الخطوات التالية:

الانحراف عن المتوسط = (س - س̄) ويرمز له بالرمز ح

$$\therefore \text{ح}_1 = (س_1 - \bar{س})$$

$$\text{ح}_2 = (س_2 - \bar{س})$$

.

.

$$ح_n = (س_n - س')$$

وبالجمع نجد أن:

$$مجد ح = مجد [(س - س') + + (س - س') + (س - س')]$$

$$= مجد س - ن س'$$

وبالتعويض عن قيمة س' (المتوسط الحسابي) بالصيغة $\frac{مجد س}{ن}$

$$\therefore مجد ح = مجد س - ن س'$$

$$= مجد س - مجد س = صفر$$

وللمتوسط الحسابي مميزات أخرى منها أن مجموع مربع الانحرافات عن المتوسط يقل عن مجموع مربع الانحرافات عن أى قيمة أخرى، وهذه خاصية هامة ستستخدم فى حساب مقاييس التشتت فيما بعد. كما أنه إذا كان لدينا عدداً من أزواج القيم لمتغيرين مستقلين فإن المتوسط الحسابي لمجموع قيم المتغيرين يساوى مجموع المتوسطين الحسابيين لهذين المتغيرين. ويمكن إثبات ذلك على النحو التالى:

$$ت_1 = أ_1 + ب_1$$

$$ت_2 = أ_2 + ب_2$$

$$\vdots$$

وبالجمع نجد أن

$$مجد ت = مجد أ + مجد ب$$

وبالقسمة على عدد المفردات ن نجد أن:

$$ت = أ + ب$$

حيث $ت$ هي المتوسط الحسابي لمجموع المتغيرين، $أ$ هي المتوسط الحسابي للمتغير الأول، $ب$ هي المتوسط الحسابي للمتغير الثاني.

وكذلك عند إضافة قيمة ثابتة «ك» إلى كل قيمة من قيم مفردات البيانات أو إذا ضربت كل قيم المفردات في قيمة ثابتة «ك»، فإن المتوسط الحسابي في الحالة الأولى يساوي المتوسط الحسابي للمفردات قبل الإضافة مضافاً إليه القيمة الثابتة بينما في الحالة الثانية يساوي المتوسط الحسابي للبيانات قبل الضرب مضروباً في القيمة الثابتة. ويمكن أن نعبر عن ذلك بالرمز كمايلي:

$$\text{متوسط (س + ك)} = \frac{\text{مجم س} + \text{ن ك}}{\text{ن}}$$

$$= \text{س} + \text{ك}$$

$$\text{ومتوسط (س} \times \text{ك)} = \frac{\text{مجم س} + \text{ن ك}}{\text{ن}}$$

$$= \text{س} \times \text{ك}$$

أما عيوب المتوسط الحسابي فتنحصر في أنه إذا احتوت مجموعة من قيم المفردات على بعض القيم الشاذة أو المتطرفة (القيم الكبيرة جداً أو الصغيرة جداً) فإن المتوسط الحسابي للمجموعة يكون في هذه الحالة مضللاً، وذلك لأن حساب المتوسط سيتأثر بهذه القيم على الرغم من قلة عددها بين مجموعة القيم، وعندئذ يفضل استخدام مقياس آخر من مقاييس النزعة المركزية لوصف هذه المجموعة. وفي التوزيعات التكرارية المتماثلة (شكل ٤-٥)، ولا يبدو هذا العيب في المتوسط الحسابي جوهرياً لعدم وجود اختلافات كبيرة بين قيم المجموعة، فنصفها يكون أكبر من المتوسط بينما النصف الآخر دونه. أما في التوزيعات الملتوية فإن الوضع يختلف عند حساب المتوسط الحسابي، إذ تؤثر القيم المتطرفة في مثل هذه التوزيعات تأثير كبيراً على قيمته. ويمكن بيان ذلك المثال التالي الذي يوضح كمية

الأمطار السنوية التى سجلها أحد المراصد فى إقليم جاف خلال عشرة أعوام متتالية (شكل رقم ٤-٩).

كمية المطر السنوى (بالبوصة) : صفر، ١، صفر، صفر، ١٠، ٢، ٢٥، صفر، صفر، ٢.

فإذا حسبنا المتوسط الحسابى للبيانات السابقة نجد أن.

$$\bar{x} = \frac{\sum x}{n} = \frac{40}{10} = 4 \text{ بوصات}$$

وبالتالى يمكن ملاحظة أن كمية الأمطار السنوية قد زادت عن متوسطها للحسابى (٤ بوصات) مرتين فقط (عندما. كلت ١٠، ٢٥ بوصة) فى السنتين المطيرتين خلال الأعوام العشرة، بينما قلت كمية الأمطار عن هذا المتوسط فى الثمانى سنوات الباقية. ومن هذا نرى كيف أن العاملين المطيرين (٣٥ بوصة أو ٨٧,٥ ٪ من كمية أمطار السنوات العشر) قد ساهما فى رفع قيمة المتوسط الحسابى وأثرا فيها أكثر مما أثرت به السنوات الأربع الأخرى التى لم تسقط بها أمطار على الإطلاق. وفى هذا المثال فإن استخدام المتوسط الحسابى كمقياس للموضع يكون مضللا حيث أن التوزيع التكرارى لهذه البيانات توزيعاً ملتوياً كما هو واضح فى الشكل رقم (٤-٩)، ويفضل فى مثل هذه الحالة استخدام مقياس آخر من مقاييس الموضع. وفى هذا المثال المحدد فإن الوسيط يساوى ٠,٥ بوصة والمنوال يساوى صفر بوصة، وكلاهما بلا شك يعطى مؤشراً أو ملخصاً للصورة العامة للنظر فى المناطق الجافة تفوق ما يعطيه المتوسط الحسابى. وبذلك يجب الحذر الشديد عند حساب المتوسط الحسابى لأى مجموعة من البيانات التى فيها قيم شديد التطرف فى الكبير والصغير إذ يعجز متوسطها الحسابى عن إعطاء الصورة المطلوبة.

ومن عيوب المتوسط الحسابى، بالإضافة إلى ما سبق، صعوبة حسابه من جداول التوزيع التكرارى غير المنتظم أو من جداول التوزيعات المفتوحة وهى

التوزيعات التكرارية غير المعروف الحد الأدنى للفقعة الأولى بها أو الحد الأعلى للفقعة الأخيرة أو كليهما معاً. ويفضل أيضاً في الحالة الأخيرة استخدام مقياس آخر من مقياس النزعة المركزية يكون أكثر ملائمة لمثل هذه الجداول المفتوحة. وأخيراً فإن المتوسط الحسابي يصعب إيجادها بالطرق البيانية متماثلاً فإن المتوسط الحسابي يقع على المحور الأفقى عند تلاقيه مع محور التماثل الرأسى بينما فى التوزيعات البسيطة الالتواء يقع المتوسط الحسابي قريباً من أكبر التكرارات من ناحية الذيل الأطوال للتوزيع والذى يلاحظ منه تأثير المتوسط بالقيم المتطرفة فتجعله يتحيز لها (شكل ٤-١٥، ب، ج).

وعلى الرغم من كل عيوب ومثالب المتوسط الحسابي السابق ذكرها فإنه كمؤشر ومقياس إحصائي له أهمية كبيرة كما يعد مفيداً بل ضرورة فى عملية البحث والتحليل الكمي فى الدراسة الجغرافية.

أما الوسيط فقد سبق أن عرفنا أنه يستخدم كقيمة متوسطة، لأنه يمثل أحياناً المتوسط المتوقع Mean Expectation أو ينوب عنه، حيث يكون عدد المفردات التى تسبقه مساوية تماماً لعدد المفردات التى تليه. كما تكون قيم المفردات لها نفس الأهمية عند حساب الوسيط بصرف النظر عن حجمها الكبيرة أو متوسطة أو صغيرة) فيما عدا القيم القيمة الوسيطة (المركزية) إذا كان عدد المفردات فردياً، أو القيمتين المتتابعتين المركزتين فى حالة إذا كانت المفردات عدداً زوجياً. ومعنى ذلك أننا قد نجد عدة مجموعات متباينة من البيانات التى تختلف بدرجة كبيرة فى مجموع قيمها ولكنها تشترك فى قيمة الوسيط. وهذا يدل على أن الوسيط الحسابية الأخرى إلا بأسلوب عام. وهو بهذا يعانى من نفس مثالب المنوال التى سيرد ذكرها بعد قليل، ومع ذلك فإن للوسيط خاصية هامة ترتبط بوضوح تعريفه ذلك لأن موقعه النسبى بين المفردات أو بين تكرار حدوثها يساعد على تحديده بسهولة. هذا بالإضافة إلى أن الوسيط يتميز بأنه أقل تأثراً بالقيم المتطرفة عن المتوسط الحسابي حيث لا تدخل هذه القيم فى حسابه، كما أنه يمكن حسابه من التوزيعات المفتوحة حين يتعذر حساب المتوسط الحسابي لها، وهو بذلك يعد

مقياساً إحصائياً مفيداً كأسلوب توضيحي يفوق المتوسط الحسابي والمنوال، وما لذلك من أهمية كبيرة في كثير من مجالات التحليل الإحصائي للبيانات الجغرافية.

على أنه في بعض الحالات يواجه حساب الوسيط بعض الصعوبات. كما في حالة المتغيرات الوثابة (غير المتصلة) ذات القيم المزدوجة. فمثلاً إذا كان لدينا عدد سكان ستة منازل بأحد الشوارع هو: ١٢١، ١٣٤، ١٥١، ١٦٢، ١٧٧، ١١٨، فإن الوسيط حسب التعريف السابق يكون:

$$\text{الوسيط} = \frac{١٦٢ + ١٥١}{٢} = \frac{٣١٣}{٢} = ١٥٦,٥$$

وهذه القيمة لا وجود لها إذ أن صفة المتغير أنه وثاب أى لا يأخذ قيمة كسرية، فلا معنى إذن لعدد سكان ١٥٦,٥ شخصاً. كما توجد صعوبة أيضاً في تحديد الوسيط في المتغيرات المتصلة ذات القيم القليلة العدد أو التي بينها قيم تتكرر بكثرة عن غيرها. فإذا كان لدينا أطوال خمسة روافد نهريّة (بالمتر) هي: ١٦٢، ١٦٢، ١٦٥، ١٦٧، فإن القيمة الوسيط هي ١٦٢ متراً ولا توجد أى قيمة أصغر منها وأن ٤٠٪ من القيم فقط أكبر منها. وبناء على ذلك فإنه كلما كان عدد مفردات البيانات صغيراً كلما كان من الأفضل استخدام مقياساً آخر غير الوسيط لوصف هذه البيانات.

ومن التعريفات السابق شرحها لمقاييس النزعة المركزية عرفنا أن المنوال بمعناه الدقيق يحدد القيم لأكثر شيوعاً وتكراراً في التوزيع، ولكن قد يصعب أحياناً تحديد موقع (قيمة) المنوال بدقة داخل التوزيع. وتظهر هذه الصعوبة في إحدى للحالتين الآتيتين: أولهما، إذا كان التوزيع يحتوى على أكثر من قيمة متساوية في تكرارها أى أكثر من منوال واحد أو قد لا يكون به منوال على الإطلاق. ففي بيانات الجدول رقم (٤-١٢) الذى يوضح انتاج الحديد الخام لبعض دول غرب أوروبا لاحظنا صعوبة تحديد منوال واضح في حالة البيانات الخاصة لكل من بلجيكا

ولكسمبرج حيث توجد فئتان من التكرارات المتساوية أو متوالان في كل مجموعة عريضة من البيانات (شكل رقم ٤-٨). وتبدو الصعوبة الثانية لتحديد قيمة المنوال في إختيار فئات التوزيع التكرارى لبيانات جدول كمية الأمطار السنوية السابق عرضه (جدول رقم ٤-١١) إلى فئات (٢٣,٩٩-٢٢), (٢٤,٩٩-٢٥), (٢٦,٩٩-٢٧) ... الخ بدلاً من (٢١-٢٢,٩٩), (٢٣-٢٤,٩٩). (٢٥-٢٦,٩٩).... الخ فيصبح جدول التوزيع التكرارى كمايلي:

الصفات (بالبوصة)	عدد مرات الحدوث (التكرار)
٢٢ - ٢٣,٩٩	١
٢٤ - ٢٥,٩٩	٩
٢٦ - ٢٧,٩٩	٣
٢٨ - ٢٩,٩٩	٨
٣٠ - ٣١,٩٩	٤
٣٢ - ٣٣,٩٩	٢
٣٤ - ٣٥,٩٩	٢
٣٦ - ٣٧,٩٩	١

وفى مثل هذا التوزيع تصبح الفئة المتوالية (٢٤ - ٢٥,٩٩) بدلاً من الفئة (٢٥ - ٢٦,٩٩) وستصبح قيمة المنوال ٢٥,١٤ بوصة بدلاً من ٢٦,٣٣ بوصة فى التوزيع السابق، كما يصبح التوزيع فى هذه الحالة مزدوج المنوال Bimodal وتعنى هذه الصعوبة أن إختيار المنوال كمقياس للموضع ليس إختياراً موفقاً، كما أنه عملياً يعد شكلاً غير دقيق للقيمة المتوسطة، وربما تنتج فى جزء منها عن الاختيار الذاتى لفئات التوزيع التكرارى كما رأينا. وعلاوة على ذلك فإن المنوال لا يحتوى على خصائص رياضية حقيقية بل له على أحسن الفروض علاقة عامة

بالمتوسط الحسابي فقط ولكن لدرجة لا تمكن من استخدامه في صيغة تشتت منها صفات أخرى للبيانات. وفيما عدا بعض الأغراض البينانية واستخدامه في بعض العمليات الحسابية العامة، فإنه لا يصلح لأن يؤخذ مقياساً للموضع في الدراسات الجغرافية كما أنه لا يوصى باستخدامه بدرجة كبيرة في التحليل الإحصائي للبيانات الجغرافية.

من كل مما سبق وحتى نتمكن من معرفة أهمية وفائدة كل مقياس من مقاييس النزعة المركزية الثلاثة، ينبغي أن نعرف نمط التوزيع التكراري الذي تلخصه هذه المقاييس في قيمة واحدة، أي معرفة الأحوال الفعلية للبيانات وكيفية انتشارها حول القيم المتوسطة. ومن الأساليب الإحصائية التي يعتمد عليها في قياس هذه الخاصية ما يعرف بمقاييس الانحراف (التشتت) والاختلاف، وهذا ما سنشرحه في الفصل التالي.

الفصل الخامس

مقاييس الإنحراف (التشتت) والاختلاف

Measures of Deviation (Dispersion) and Variability

مقاييس الانحراف (التشتت) والاختلاف

تكلمنا في الفصل السابق عن المقاييس المختلفة للنزعة المركزية (المتوسطات) وكيفية قياس كل منها، وعرفنا أنها تهدف إلى تبسيط وتحديد الاتجاه العام لمجموعة من قيم مفردات البيانات عن طريق تلخيصها في قيمة متوسطة مركزية واحد. تصف أو تمثل قيم هذه المجموعة وهي ما أطلقنا عليه صفة «التركز» في البيانات. وكما لاحظنا فإن مقاييس النزعة ركزية لا تكفى وحدها لإعطاء فكرة دقيقة عن مجموعة من البيانات وصفها وصفاً كاملاً إذ أنها لا تبين طبيعة هذه البيانات ولا كيفية توزيع مفرداتها، كما أن استخدامها فقط لمقارنة عدة مجموعات من البيانات لا يكفى لإظهار حقيقة المقارنة. وعلى ذلك فالوصف الدقيق لمجموعة من القيم الأصلية للبيانات أو مقارنة هذه المجموعة بمجموعة أو عدة مجموعات أخرى بدقة يجب ألا يقتصر على وصف أو مقارنة متوسطاتها، بل يجب الوقوف على صفات كل مجموعة. ويتحقق ذلك بوصف درجة الاختلاف بين قيم مفردات المجموعة وتباعدها عن قيمتها المتوسطة، أو بعبارة أخرى وصف درجة إنحراف أو تشتت هذه القيم.

والمقصود إحصائياً بالانحراف أو التشتت هو مدى تباعد وتناثر (إنتشار - Scat-ter) قيم مفردات البيانات عن بعضها البعض. فإذا كانت قيم المفردات متقاربة من بعضها البعض (أو إذا تساوت جميع القيم) فإن مدى التناثر يكون صغيراً وبالتالي

يدل على تجانس هذه القيم، أما إذا كانت القيم منتشرة فيما بينها أى متباعدة عن بعضها البعض فإن مدى التناثر يكون كبيراً، ويتخذ ذلك دليلاً على عدم التجانس بين قيم المفردات. وعلى ذلك يمكننا إتخاذ مقدار إنحراف أو تشتت القيم كمقياس لتجانس البيانات أو كمقياس لتركز القيم وقربها من بعضها أو لتبعثرها وتباعدها بعضها عن البعض. وتقاس درجة التجانس أو التركيز ببعض المقاييس التي تعرف باسم مقاييس الإنحراف (التشتت) والاختلاف، بينما تقاس اتجاهات التركيز والتطرف في القيم بما يسمى بمقاييس الالتواء Skewness ومقاييس التفرطح Kurtosis.

أنواع الإنحراف:

إذا أردنا توضيح شكل التشتت لمفردات المتغيرات موضع الدراسة بيانياً، فإن أبسط وأنسب وسيلة لتحقيق ذلك هو ما يتم بواسطة منحى التوزيع التكرارى Fre-quency Distribution Curve الذى سبق شرح وتطبيق طريقة تمثيله، بينما «شكل الانتشار Scatter Diagram» الذى تتحرك فيه قيم المتغيرات لايفيد كثيراً فى بيان شكل التشتت بقدر ما يفيد فى توضيح نوع ودرجة العلاقة الارتباطية بين المتغيرات (سبق شرح هذا النوع من الرسوم عند دراسة طرق العرض البيانى للبيانات الاحصائية).

أما إذا أردنا التعبير عن التشتت لمجموعة من قيم المفردات بصورة رقمية فإن ذلك يتم عن طريق مقارنة أقل القيم بأكبرها فى قيم المجموعة الواحدة أو بما يعرف بالمدى المطلق Range. فعلى سبيل المثال، إذا كانت لدينا عينتين لأطوال خمسة من الأودية النهرية (بالكيلو متر) فى منطقتين مختلفتين تتوزع قيمها على النحو التالى:

العينة الأولى: ١٧، ١٤، ١٠، ٧، ٢ (المتوسط = ١٠، الوسيط = ١٠)

العينة الثانية: ١٢، ١١، ١٠، ٩، ٨ (المتوسط = ١٠، الوسيط = ١٠)

فنلاحظ أن قيمة المتوسط الحسابي والوسيط لكل من هاتين العينتين متساوية أى أن العينتين تشتركان فى أكثر من مقياس من مقاييس الموضع ومع ذلك نجد أن هناك فرقاً كبيراً بين العينتين من حيث طبيعة توزيع قيم المفردات فى كل منهما، إذ تتوزع أطوال أودية العينة الأولى فى مدى تبعثر أكبر (من ٢ - ١٧ كيلو متراً) بينما تتوزع أطوال أودية العينة الثانية فى مدى تبعثر أقل (من ٨ - ١٢ كيلو متراً). ومعنى ذلك أن التشتت بين مفردات العينة الأولى أكبر منه بين مفردات العينة الثانية (أى أن قيم العينة الأولى أقل تجانساً من العينة الثانية). وكذلك نلاحظ أن هناك إختلاف واضح فى انحراف قيم المفردات عن المتوسط الحسابى فى كل عينة إذ ينحصر الانحراف عن المتوسط للعينة الأولى بين +٧ و -٨ بينما ينحصر فى العينة الثانية بين +٢ و -٢.

ومهما يكن من أمر فإن المدى كمقياس للتشتت، رغم بساطته، لا يعد دقيقاً للتعبير من وصف مجموعة من قيم المفردات بالتجانس (التشتت) فهو لا يعطى فكرة جيدة عن مدى تباعد القيم عن بعضها البعض، وربما يكون مضللاً فى الحالات التى يوجد فيها قيمة متطرفة (شاذة) فى البيانات مما يسبب زيادة كبيرة فى المدى الذى يستدل منه على أن قيم المفردات غير متجانسة ومشتتة تشتتاً كبيراً بينما تكون القيم كلها - ما عدا القيمة المتطرفة - متقاربة. كما أن المدى لا يعطى ملخصاً كافياً للتشتت أو التبعثر فى قيم المفردات، أو لايوجز التشتت فى قيمة واحدة يعتمد عليها فى التحليل الإحصائى للبيانات. وبناء على ذلك هناك ثلاثة مقاييس أخرى أكثر دقة وفائدة من حيث تلخيص التشتت وإنتشار ومفردات البيانات ذلك لأنها تهدف إلى، أو ينتج عنها، قيم رقمية تعرف باسم «القيم الانحرافية Deviation Values»، وهى عبارة عن إختلافات أو فروق قيم مفردات المجموعة عن المتوسط. وهذه المقاييس هى: الانحراف الربيعى، الانحراف المتوسط، للتباين (والانحراف المعيارى). والمقياس الأول من هذه المقاييس يستخدم الوسيط، بينما يستخدم المتوسط الحسابى فى المقياسين الآخرين. أما المتوال فلا يستخدم فى

حالة قياس التشتت كبديل للمتوسط لأنه لا يظهر القيم المنحرفة بشكل فعال، وهذه ولاشك واحدة أخرى من مساوئ المنوال التي تقلل بدرجة كبيرة جداً إستخدامه فى التحليل الجغرافى الكمى. وفيما يلى شرح للمقاييس الثلاثة كل على حدة وبيان حساب التشتت بواسطتها.

١ - الانحراف الربيعى Quartile Deviation

يستخدم الانحراف الربيعى أو ما يسمى بنصف المدى الربيعى - Semi Inter quartile Range لقياس التشتت فى البيانات التى تتميز بتطرف بعض قيم مفرداتها، ويهدف إلى التغلب على أهم عيوب المدى المطلق - وهو إعتماده على القيم المتطرفة، وذلك بحذف الربع الأول والربع الأخير من قيم المتغير موضع الدراسة وحساب المدى للقيم الأخرى. أو بعبارة أخرى يحدد النصف الأوسط لمجموعة قيم المتغير بعد ترتيبها حسب قيمتها ويستخرج المدى الذى تقع فيه القيم الوسيطة، وكلما كان المدى الذى ينتشر عليه هذا النصف من القيم كبيراً كان التشتت كبيراً بين قيم المتغير والعكس يكون صحيحاً، وبذلك يمكن الحصول على مقياس للتشتت يعتبر أفضل من مقياس المدى المطلق.

وتجدر الإشارة إلى أن طريقة إيجاد الانحراف الربيعى لمجموعة من القيم تشبه الطريقة التى يوجد بها الوسيط، ولذلك فإن مزايا ومثالب كل منهما واحدة. فالوسيط كما سبق أن عرفنا هو القيمة التى تقسم قيم المجموعة المرتبة حسب قيمتها (تصاعدياً أو تنازلياً) إلى نصفين متساويين فى العدد. ويمكن قياس التشتت عن القيمة الوسيطة بقسمة كل نصف المجموعة إلى نصفين آخرين، وبذلك تصبح المجموعة مقسمة إلى أربعة أقسام متساوية - ويسمى كل حد من حدود التقسيم «بالربيع»، ويكون عدد الحدود الفاصلة فى هذه الحالة ثلاثة: يطلق على الحدين الفاصلين الأول والأخير بالربيع الأدنى Lower Quartile والربيع الأعلى Upper Quartile على الترتيب، بينما الحد الثالث (الأوسط) هو الوسيط. ويظهر ذلك فى الشكل التالى (شكل رقم: ١-٥) الذى يمثل بيانات كمية المطر السنوى التى

النهاية العظمى	٣٦,٥٠	١
	٣٤,٨١	٢
	٣٤,٤٢	٣
	٣٣,٣٤	٤
	٣٢,٨٧	٥
	٣١,٩٣	٦
	٣٠,٩٢	٧
الربيع الأعلى	٣٠,٩٥	٨
	٣٠,١٧	٩
	٢٩,١٢	١٠
	٢٩,١١	١١
	٢٨,٩٥	١٢
	٢٨,٩٠	١٣
	٢٨,٥٩	١٤
الوسيط (٢٨,٢٧)	٢٨,٤٥	١٥
	٢٨,٠٨	١٦
	٢٨,٠٠	١٧
	٢٦,٨٣	١٨
	٢٦,٥٧	١٩
	٢٦,٠٢	٢٠
	٢٥,٩٧	٢١
	٢٥,٣٨	٢٢
الربيع الأدنى	٢٥,٥٧	٢٣
	٢٥,٢٧	٢٤
	٢٥,١٩	٢٥
	٢٥,١٨	٢٦
	٢٥,١٥	٢٧
	٢٤,٨٧	٢٨
	٢٤,٠١	٢٩
النهاية الصغرى	٢٢,٤٧	٣٠

شكل رقم (١-٥)

الوسيط والربيعين (الأدنى والأعلى) لكميات المطر السنوية المسجلة بمرصد Bidston -
المجلترا في الفترة من ١٩٠١ - ١٩٣٠

سجلها مرصد Bidston - إنجلترا في الفترة من ١٩٠١ - ١٩٣٠ (جدول رقم ٤ - ١٠).

من الشكل رقم (٥-١) يتضح أن الربيع الأدنى يفصل ٢٥٪ من المفردات التي تشمل أقل القيم عن بقية المفردات، بينما يفصل الربيع الأعلى ٢٥٪ من المفردات التي تحتوى على أعلى القيم. أى أن حدودهما تحصر القيم الوسطى (أو ٥٠٪ من المفردات) التي تشتت حول قيمة الوسيط. ومن الشكل أيضاً نرى أن قيمة الوسيط ٢٧, ٢٨ بوصة، وهى نفس القيمة التي حصلنا عليها سابقاً عند تطبيق طريقة حساب الوسيط لهذا المثال، أما قيمة الربيع الأدنى فهى عبارة عن القيمة الوسيطة لنصف عدد القيم أسفل الوسيط، أى أنها القيمة الثامنة (من أسفل) من بين الخمس عشرة قيمة التي تقع أسفل الوسيط والتي تساوى ١٥, ٥٧ بوصة. وبنفس الطريقة نحصل على قيمة الربيع الأعلى وهى القيمة الثامنة (من أعلى) من بين الخمس عشرة قيمة التي تقع فوق الوسيط والتي تساوى ٣٠, ٥٩ بوصة. ويعرف الفرق بين قيمتى الربيع الأعلى والربيع الأدنى وهو ٣٠, ٥٩ - ١٥, ٥٧ = ١٥, ٠٢ بوصة بالمدى الربيعى. وهذا المدى يقع منحرفاً عن الوسيط، ولكن فى الحالات التي يكون فيها منحنى التوزيع التكرارى للبيانات متماثلاً ومتوازناً تماماً (منحنى معتدل) فإن كلاً من الربيعين سيقع فى منتصف المسافة بين الوسيط والقيم النهائية (العليا والسفلى) للتوزيع. أى أن كلاً منهما سيبعد عن الوسيط بنصف قيمة المدى وهى ١٥, ٠٢ ÷ ٢ = ٧, ٥١ بوصة للمثال بين أيدينا. وهذه القيمة تعرف بالإنحراف الربيعى أو نصف المدى الربيع وهى التي تتخذ كدليل أو مؤشر يبين مقدار مدى قيم النصف الأوسط حول الوسيط لمجموعة من المفردات المرتبة حسب قيمتها. ويمكن أن نعبر عن ذلك بالصيغة الآتية:

$$\text{الانحراف الربيعى} = \frac{\text{قيمة الربيع الأعلى} - \text{قيمة الربيع الأدنى}}{2} \dots (٥ - ١)$$

ايجاد الانحراف الربيعى للبيانات المبوبة:

لحساب الانحراف الربيعى من بيانات مبوبة أى من جداول التوزيعات التكرارية تتبع الخطوات التالية:

١- تحسب من جدول التوزيع التكرارى التوزيع التكرارى المتجمع الصاعد للتكرارات.

٢- توجد ترتيب كل من الربيعين الأدنى $(\frac{n}{4})$ والأعلى $(\frac{3 \times n}{4})$.

٣- تعين قيمة كل من الربيعين باستخدام الصيغة الآتية:

$$\text{قيمة الربيع} = \text{الحد الأدنى للفئة الربيعية} + \wedge$$

$$\left(\frac{\text{ترتيب الربيع} - \text{تكرار المتجمع السابق للفئة الربيعية}}{\text{التكرار الأصلي للفئة الربيعية}} \right) \times \text{طول الفئة}$$

ويمكن بيان ذلك المثال السابق الخاص بكميات المطر السنوية المسجلة بمرصد Bidston إنجلترا فى الفترة من ١٩٠١ - ١٩٣٠.

جدول رقم (١-٥)

كمية الأمطار السنوية المسجلة بمرصد Bidston

المجلترا فى الفترة من ١٩٠١ - ١٩٣٠

التكرار المتجمع الصاعد		التكرار	الفئات (بوصة)
التكرار المتجمع الصاعد	الحدود العليا للفئات		
١	أقل من ٢٣	١	٢٢,٩٩ - ٢١
٣	أقل من ٢٥	٢	٢٤,٩٩ - ٢٣
١٣	أقل من ٢٧	١٠	٢٦,٩٩ - ٢٥
١٩	أقل من ٢٩	٦	٢٨,٩٩ - ٢٧
٢٤	أقل من ٣١	٥	٣٠,٩٩ - ٢٩
٢٦	أقل من ٣٣	٢	٣٢,٩٩ - ٣١
٢٩	أقل من ٣٥	٣	٣٤,٩٩ - ٣٣
٣٠	أقل من ٣٧	١	٣٦,٩٩ - ٣٥
		٣٠	المجموع

ويكون ترتيب الربيع الأدنى $٧,٥ = \frac{٣٠}{٤}$

∴ قيمة الربيع الأدنى $٢ \times \left(\frac{٣-٧,٥}{١٠} \right) + ٢٥٠٠ =$

$$\frac{٢ \times ٤,٥}{١٠} + ٢٥,٠ =$$

$$٢٥,٩ = ٩ + ٢٥,٠ =$$

$$\text{وترتيب الربيع الأعلى} = \frac{3 \times 30}{4} = 22,5$$

$$\therefore \text{قيمة الربيع الأعلى} = 29,0 + \left(\frac{19 - 22,5}{5} \right) \times 2$$

$$= \frac{2 \times 3,5}{5} + 29,0 =$$

$$= 29,0 + 1,4 = 30,4 \text{ بوصة}$$

$$\text{يكون الانحراف الربيعي أو (نصف المدى الربيعي)} = \frac{\text{قيمة الربيع الأعلى} - \text{قيمة الربيع الأدنى}}{2}$$

$$= \frac{30,4 - 25,9}{2} =$$

$$= \frac{4,5}{2} = 2,25 \text{ بوصة}$$

ومن الواضح أن نصف عدد التكرارات (سنوات المطر) ينحصر بين قيمتي يعين الأدنى والأعلى (أى الكميتين 25,9 و 30,4 بوصة)، وهذا النصف وسط الذى لا يتأثر بالقيم المتطرفة، كما أنه يمثل 50% من البيانات التى تشتت بها عن الوسيط بمقدار \pm نصف المدى الربيعي. وبناء على ذلك يمكن وصف حراف الربيعي بأنه معدل الانحراف المتوقع عن القيمة الوسيطة أو بمعنى آخر نصف القيم تختلف عن الوسيط بأكثر من مقدار نصف المدى الربيعي، بينما يف الآخر يختلف عنه بأقل من هذا المقدار.

ومن الملاحظ أيضاً أنه إذا قارنا الانحراف الربيعي (2,25) وقيمة المدى المطلق $21 - 3 = 18$ بوصة) وهما يقيسان التشتت لنفس سنوات المطر - نجد أن تهما مختلفتان، وهذا أمر متوقع فكل منهما يقوم على أساس مختلف. ولذلك يتحتم عند إجراء مقارنة بين تشتت عدة مجموعات من البيانات أن يكون التشتت بطريقة واحدة لكل المجموعات حتى تكون المقارنة صحيحة. وعلى

أية حال، فإن الانحراف الربيعي أو نصف المدى الربيعي لا يعد مقياساً دقيقاً لتشتت البيانات بل يعتبر مقياساً لتوضيح مدى تشتت البيانات بعداً أو قريباً من مقياس الموضوع (الوسيط). ومع ذلك فهو وسيلة هامة لبيان مدى تشتت البيانات حول نقطة معينة، ولذلك فهو يستخدم كثيراً في تحليل البيانات المناخية، أو لإبراز المعلومات الاقتصادية وتحليلها - في مجال الجغرافية الاقتصادية - وبصفة خاصة عند دراسة وتحليل التوطن الصناعي في أحد الأقاليم الجغرافية.

الانحراف المتوسط : Mean Deviation

لاحظنا من حساب كل من المدى المطلق والانحراف الربيعي أن كلا منهما لا يعطى صورة كاملة عن التشتت في قيم البيانات. فالمدى يعتمد على القيم الطرفية المحددة للتوزيع، التي غالباً ما تكون متطرفة، ولا يهتم بتوزيع بقية القيم، بينما يهمل الانحراف الربيعي نصف عدد للقيم في أطراف توزيع البيانات. وكما سبق القول أن القياس الصحيح للتشتت يقوم على أساس مدى تباعد القيم عن بعضها فإذا كانت القيم قريبة من بعضها فإنه تكون متجمعة (مركزة) حول قيمة في الوسط وكلما كانت متناثرة (مبعثرة) كلما تباعدت عن هذه القيمة. وبناء على ذلك فإننا نلجأ إلى قياس التشتت على أساس بين القيم المختلفة للبيانات وقيمة متوسطة مثل المتوسط الحسابي (الذي يعتبر أكثر مقاييس النزعة المركزية حساسية للتغيرات في أى قيمة من القيم).

ولما كان من المعروف أن مجموع انحرافات القيم عن متوسطها الحسابي يساوى صفراً وذلك لأن مجموع الانحرافات الموجبة عن المتوسط الحسابي يساوى مجموع الانحرافات السالبة عنه، ولذا فإنه لا يمكن أخذ مجموع الانحرافات عن المتوسط الحسابي كمقياس للتشتت، ولكن إذا أهملت إشارات الانحرافات (انحراف موجب أو انحراف سالبة) وجمعنا الانحرافات المجردة من الإشارة ثم قسمنا هذا المجموع على عدد القيم فإننا نحصل بذلك على مقياس للتشتت يطلق عليه اسم الانحراف المتوسط (شكل رقم ٥-٢). ويمكن تلخيص ذلك في الصيغة الجبرية الآتية:



$$\frac{\text{مجا ح ا}}{ن} = \text{الانحراف المتوسط} \quad \dots \dots \dots (5-2)$$

حيث مجا هي المجموع، ن هي عدد القيم، ح هي الانحراف عن المتوسط الحسابي، ويرمز الخطان الرأسيان إلى أن الانحرافات تؤخذ بدون إشارات الجبرية (أي الانحرافات المطلقة). وتمتاز قيمة الانحراف المتوسط بأنها تكبر كلما ازداد تشتت أو التفاوت بين القيم وبعدت عن متوسط المجموعة، بينما تقل كلما قل تشتت وانتشار القيم. كما يمتاز الانحراف المتوسط بأنه يأخذ في الاعتبار جميع قيم المفردات المتاحة.

ويتم قياس التشتت لمجموعة صغيرة من البيانات بواسطة الانحراف المتوسط بأن نوجد أولاً المتوسط الحسابي للقيم (يأخذى الطرق السابق شرحها) ثم نحسب الانحراف (الفرق) بين هذا المتوسط وبين كل قيمة من قيم البيانات ونجمع هذه الانحرافات - مع استبعاد الاشارات - ثم نقسمها على عدد القيم، والمثال التالي يوضح ذلك.

مثال. في مجموعة البيانات الخاصة بأطوال خمسة من الأودية النهرية في منطقة ما (بالكيلو متر) ١٧، ١٤، ١٠، ٧، ٢ يكون المتوسط الحسابي $\frac{50}{5} = 10$ كيلو متراً، وتكون انحرافات القيم عن هذا المتوسط هي ٧ +، ٤ +، صفر، ٣-، ٨- وبإهمال الإشارة السالبة يكون الانحراف المتوسط هو $\frac{8+3+0+4+7}{5} = \frac{22}{5} = 4,4$ كيلو متراً.

ولإيجاد الانحراف المتوسط لتوزيع تكرارى نحسب المتوسط الحسابي بالطرق العادية ثم نحسب انحرافات مراكز الفئات عن المتوسط الحسابي مع إهمال الإشارة (ح ا)، ثم نوجد حاصل ضرب تكرار كل فئة في انحراف مركز الفئة مع إهمال الإشارة (ح ا × ك)، ونجمع هذه الانحرافات ونقسم الناتج على عدد التكرارات، وبذلك يكون:

$$\text{الانحراف المتوسط} = \frac{\text{مجا ح ك}}{\text{مجا ك}} \dots \dots \dots (3-5)$$

والجدول الآتي يبين خطوات حساب الانحراف المتوسط لتوزيع القرى في أحد المراكز حسب الحجم السكاني وذلك على أساس أن المتوسط الحسابي للحجم السكاني للقرى هو ١٩٧٦٤ نسمة.

جدول رقم (٥ - ٢)

الانحراف المتوسط لتوزيع القرى في أحد المراكز
حسب الحجم السكاني لهذه القرى

فئات الحجم (نسمة)	التكرار ك	مراكز الفئات م	الانحراف عن المتوسط ح ك	ح ك × ك
١ - ٥٠٠٠	١٢	٢٥٠٠,٥	١٧٢٦٣,٥	٢٠٧٢٦٢,٠
٥٠٠١ - ١٠٠٠٠	١٧	٧٥٠٠,٥	١٢٢٦٣,٥	٢٠٨٤٧٩,٥
١٠٠٠١ - ١٥٠٠٠	١٩	١٢٥٠٠,٥	٧٢٦٣,٥	١٤٨٠٠٦,٥
١٥٠٠١ - ٢٠٠٠٠	١٤	١٧٥٠٠,٥	٩٢٦٣,٥	٣١٦٨٩,٠
٢٠٠٠١ - ٢٥٠٠٠	٧	٢٥٥٠٠,٥	٢٧٣,٥	١٩١٤٨,٥
٢٥٠٠١ - ٣٠٠٠٠	٥	٢٧٥٠٠,٥	٧٧٣,٥	٣٩٦٧٧,٥
٣٠٠٠١ - ٣٥٠٠٠	٥	٣٢٥٠٠,٥	١٢٧٣,٥	٦٣٦٧٧,٥
٣٥٠٠١ - ٤٠٠٠٠	٢	٣٧٥٠٠,٥	١٧٧٣,٥	٣٥٤٧١,٠
٤٠٠٠١ - ٤٥٠٠٠	٥	٤٢٥٠٠,٥	٢٢٧٣,٥	١١٣٦٧٧,٥
٤٥٠٠١ - ٥٠٠٠٠	٢	٤٧٥٠٠,٥	٢٧٧٣,٥	٥٥٤٧١,٠
٥٠٠٠١ - ٥٥٠٠٠	٣	٥٢٥٠٠,٥	٣٢٧٣,٥	٩٨٢٠٦,٥
٥٥٠٠١ - ٦٠٠٠٠	٢	٥٧٥٠٠,٥	٣٧٧٣,٥	٧٥٤٧١,٠
٦٠٠٠١ - ٦٥٠٠٠	٢	٦٢٥٠٠,٥	٤٢٧٣,٥	٨٥٤٧١,٠
المجموع	٩٥			١١٨١٧٠٨,٥

$$\text{فيكون الانحراف المتوسط} = \frac{١١٨١٧٠٨,٥}{٩٥} = ١٢٤٣٩ \text{ نسمة تقريباً.}$$

وتجدر الإشارة إلى أن الانحراف المتوسط لا يستخدم إلا قليلاً كمقياس للتشتت عند التحليل الكمي للبيانات الجغرافية، ولكننا درسناه هنا كمقدمة لمقياسين آخرين من أكثر مقاييس التشتت إنتشاراً ودقة وهما التباين والانحراف المعياري.

٣- التباين والانحراف المعياري: Variance and Standard Deviation

يعد التباين والانحراف المعياري من أكثر مقاييس التشتت استخداماً في مختلف البحوث والدراسات في العلوم الأرضية والاجتماعية، لما لهما من أهمية كبيرة في الدراسات التحليلية ولا اعتماد كثير من المقاييس الاحصائية عليهما. وهما مثل الانحراف المتوسط يعتمدان على استخدام انحرافات القيم عن المتوسط الحسابي. ولكن بدلاً من إهمال الإشارات وللتغلب على الانحرافات السالبة، كما في حالة الانحراف المتوسط، يتم تربيع الانحرافات فتصبح كلها موجبة ثم يقسم مجموع مربع الانحرافات على عدد هذه الانحرافات فنحصل على متوسط مربع الانحرافات (الانحراف التربيعي المتوسط) أو ما يعرف باسم التباين Variance ويرمز له بالرمز (σ^2) في حالة المجتمع والرمز (s^2) في حالة العينة، ويمكن وضع ذلك في الصيغة الجبرية التالية:

$$\text{تباين المجتمع } (\sigma^2) = \frac{\sum (x_i - \mu)^2}{n} \quad \dots \dots \dots (5 - 4)$$

حيث μ هي القيمة المعطاة، μ تشير إلى المتوسط الحسابي للمجتمع و n هي عدد القيم. ويكون تباين العينة هو:

$$\text{تباين المجتمع } (s^2) = \frac{\sum (x_i - \bar{x})^2}{n} \quad \dots \dots \dots (5 - 5)$$

حيث \bar{x} تشير إلى المتوسط الحسابي للعينة. وحيث أن قيمة المتوسط الحسابي للمجتمع (μ) نادراً ما تكون معروفة فإن تباين العينة يتخذ كـتقدير غير متحيز لتباين المجتمع - ويتم ذلك بتصحيح تباين العينة أى بضربه في عامل التصحيح أو تصحيح

«بسل» Besel's Correction $\frac{n}{n-1}$ ، ويعرف الناتج «بأحسن تقدير a best estimate

لتباين المجتمع ويرمز له بالرمز $(\hat{\sigma}^2)$ ويمكن التعبير عن ذلك بالصيغة الآتية:

$$\begin{aligned} \hat{e}^2 &= \text{تباين العينة} \times \frac{n}{1-n} \\ &= \frac{n}{1-n} \times \frac{\text{مجم} (s - \bar{s})^2}{n} \\ &= \frac{\text{مجم} (s - \bar{s})^2}{1-n} \end{aligned}$$

(٥ - ٦)

أى تباين العينة يساوى مجموع مربع الانحرافات عن المتوسط الحسابى للعينة (س) مقسوماً على عدد الانحرافات المستقلة (ن - ١). وقيمة التباين لا تقل عن الصفر ولكنها تساوى صفراً فى حالة ما إذا كانت جميع قيم مفردات العينة متساوية. وكلما زادت قيمة التباين دل ذلك على زيادة التشتت والاختلاف بين قيم المفردات أو تبعثرها حول المتوسط الحسابى وبالتالى على قلة تجانسها. أما إذا قلت قيمة التباين كان ذلك دليلاً على تركيز (تجمع) قيم المفردات حول المتوسط الحسابى وبالتالى زيادة تجانسها. أى أن أمدى تجانس مفردات المجموعة يتناسب عكسياً مع قيمة التباين لها.

مثال:

إذا كانت لدينا البيانات الآتية التى تمثل ثلاث عينات تحتوى كل منها على أعمار ستة أشخاص وأردنا حساب تباين كل مجموعة فستكون خطوات الحساب على النحو المبين بالجدول رقم (٥ - ٣).

وبمقارنة التباين بالانحراف المتوسط - كمقياس للتشتت - للمجموعات السابقة من البيانات والتى تمثل كل مجموعة منها نمطاً مختلفاً للانحرافات عن المتوسط (شكل رقم : ٥ - ٣) نلاحظ أنه على الرغم من أن قيمة الانحراف المتوسط للمجموعات الثلاث واحدة، إلا أنه يمكن القول بأن بيانات المجموعة (أ) تتسم بأنها أعظم تبايناً وانتشاراً حول المتوسط من بيانات المجموعة (ب) والمجموعة (ج) بينما تتخذ وضعاً متوسطاً بين المجموعتين الأخريتين. والوسيلة التى يمكن بها تأكيد الانحرافات الكبيرة على حساب الانحرافات الصغيرة هو ما يتم بواسطة تربيع كل الانحرافات ويظهر ذلك جلياً فى بيانات المجموعة (ج) فى الشكل السابق.

ويعتبر التباين من أشهر مقاييس التشتت (الاختلاف) إلا أنه يقيس مربع الانحرافات. وللحصول على مقياس للتشتت محسوباً بنفس وحدات القيم الأصلية فإننا نستخرج الجذر التربيعي للتباين، وتعرف القيمة الناتجة بالانحراف المعياري. ويرجع السبب في أخذ الجذر التربيعي للحصول على هذه القيمة إلى تربيع الانحرافات في البداية. ولكي تعود الوحدات إلى قيمتها الأصلية بعد التربيع لابد من أخذ الجذر التربيعي ليكون التشتت مقيساً بنفس وحدات القيم الأصلية (بوصات، أمتار، أطنان، سنوات ... إلخ). وبناء على ذلك فإن الانحراف المعياري للمجتمع يكون:

$$\sigma = \sqrt{\frac{\text{مجمد } (م - \bar{م})^2}{n}} \quad \dots \dots \dots (٧ - ٥)$$

ويكون الانحراف المعياري للعينة هو

$$\sigma_e = \sqrt{\frac{\text{مجمد } (م - \bar{م})^2}{n}} \quad \dots \dots \dots (٨ - ٥)$$

مثال:

لحساب قيمة الانحراف المعياري من مجموعة أو عدة مجموعات من البيانات (مثل حساب تباين أعمار ستة أشخاص للثلاث عينات السابق الإشارة إليها) نأخذ الجذر التربيعي لقيم تباين كل مجموعة. وإذا استخدمنا نفس بيانات الجدول رقم (٣-٥) فإن الانحراف المعياري لكل عينة يكون كالآتي:

$$\sigma_1 = \sqrt{1} = 1 \text{ سنة}$$

$$\sigma_2 = \sqrt{1,67} = 1,29 \text{ سنة}$$

$$\sigma_3 = \sqrt{3} = 1,73 \text{ سنة}$$

وبالمثل يحسب الانحراف المعياري للعينة كتقدير غير متحيز للانحراف المعياري

للمجتمع الذى أخذت منه، أو ما يسمى أيضاً «أحسن تقدير للانحراف المعيارى للمجتمع (ع)» - وللحصول على هذا التقدير غير المتحيز للانحراف المعيارى للمجتمع فإن الانحراف المعيارى للعينة يصبح بعامل «بسل» وذلك على النحو التالى:

$$\hat{e} = \frac{n}{1-n} \sqrt{\text{الانحراف المعيارى للعينة}} \times \sqrt{\frac{\text{مجد (م-م')}}{n}} = \dots \dots (9-5)$$

وللتبسيط والاختصار توضع الصيغة (9-5) على النحو التالى:

$$e = \sqrt{\frac{\text{مجد (م-م')}}{n}} = \dots \dots (10-5)$$

ومما تجدر الإشارة إليه أن التباين والانحراف المعيارى - كمقياسين كميين للاختلاف أو التبعر (للتشتت) - يختلفان فى مجال الاستفادة منهما فى التحليلات الجغرافية، فالمقياس الأول (التباين) يصلح لأغراض التنبؤ والتكهن ويستعمل بكثرة فى الاختبارات الاحصائية، بينما يصلح الانحراف المعيارى للأغراض الوصفية فقط حيث أنه يقاس بنفس وحدات الصفة المتغيرة. أمثلة نوعية:

بعد ذلك العرض الموجز لطريقة حساب كل من التباين والانحراف المعيارى نحاول الآن تطبيقهما على بيانات نوعية فى مجال الدراسات الجغرافية حتى نتعرف على مجال الاستفادة منهما فى التحليل الكمي الجغرافى. فالجدول رقم (5-4) يوضح كمية الأمطار السنوية التى سجلها مرصد Bidston - إنجلترا (السابق الإشارة إليها) وذلك فى الفترة من ١٩٠١ إلى ١٩٣٠، وبحساب كل من الانحراف المتوسط، التباين والانحراف المعيارى بالطرق السابق ذكرها (المعادلات ٥

(٨ -) وجد أن قيمة كل من الانحراف المتوسط والانحراف المعياري هي على الترتيب ٢,٧٩ بوصة و ٣,٤٥ بوصة. ومن الواضح أنه إذا قارنا الانحراف المتوسط (٢,٧٩ بوصة) وقيمة الانحراف المعياري (٣,٤٥ بوصة) وهما يقيسان التبعر والتشتت لنفس المجموعة من كميات المطر - نجد أن قيمتهما مختلفتان وهذا أمر متوقع لأنه على الرغم من أنهما يقيسان نفس درجة الاختلاف بين قيم مجموعة واحدة من البيانات إلا أن كلاً منهما يحسب بطريقة مختلفة عن الآخر. كذلك فإنه يمكن استنتاج علاقة اعتبارية تربط بين الانحراف المعياري والانحراف المتوسط للتوزيعات المتوسطة الالتواء والتي يعبر عنها بالصيغة: الانحراف المتوسط = $\frac{1}{2} \times$ الانحراف المعياري، أو بمعنى آخر أن الانحراف المعياري أكبر من الانحراف المتوسط بنسبة ٢٥٪، أي أن الانحراف المعياري = $1,٢٥ \times$ الانحراف المتوسط، ويظهر ذلك جلياً من النتائج التي حصلنا منه الجدول رقم (٥ - ٤) والتي توضح أن العامل الذي يرجع به الانحراف المتوسط يجب أن يكون ١,٢٤ ليعطى نفس الانحراف المعياري الناتج بالمعادلة (٥ - ٨).

جدول رقم (٥ - ٤): حساب الانحراف المتوسط والانحراف المعياري
لكمية المطر السنوي في مرصد Bidston - إنجلترا
في الفترة من ١٩٠١ - ١٩٣٠

السنة	القيم	ح	ح
١٩٠١	٢٥,١٩	٣,٢٦-	١٠,٦
١٩٠٢	٢٥,٥٧	٢,٨٨-	٨,٣
١٩٠٣	٣٤,٤٢	٥,٩٧+	٣٥,٦
١٩٠٤	٢٥,١٨	٣,٢٧-	١٠,٧
١٩٠٥	٢٤,٠١	٤,٤٤-	١٩,٦
١٩٠٦	٢٨,٠٨	٠,٣٧-	٠,١
١٩٠٧	٢٦,٥٧	١,٨٨-	٣,٥
١٩٠٨	٢٨,٩٠	٠,٤٥+	٠,٢
١٩٠٩	٢٨,٤٥	صفر	صفر

٠,٠٢	,١٤+	٢٨,٥٩	١٩١٠
١٠,١	٣,١٨-	٢٥,٢٧	١٩١١
٢,٩٥	١,٧٢+	٣٠,١٧	١٩١٢
٧,١	٢,٦٧-	٢٥,٧٨	١٩١٣
٥,٩	٢,٤٣-	٢٦,٠٢	١٩١٤
٢,٦	١,٦٢-	٢٦,٨٣	١٩١٥
١٢,٨	٣,٥٨-	٢٤,٨٧	١٩١٦
٤,٦	٢,١٤+	٣٠,٥٩	١٩١٧
١٢,١	٣,٤٨+	٣١,٩٣	١٩١٨
٠,٤٥	٠,٦٧+	٢٩,١٢	١٩١٩
٢٣,٩	٤,٨٩+	٣٣,٤٣	١٩٢٠
٣٥,٧	٥,٩٨-	٢٢,٤٧	١٩٢١
٦,١٥	٢,٤٨-	٢٥,٩٧	١٩٢٢
٦,١	٢,٤٧+	٣٠,٩٢	١٩٢٣
١٩,٦	٤,٤٢+	٣٢,٨٧	١٩٢٤
٠,٢	٠,٤٥-	٢٨,٠٠	١٩٢٥
٠,٢٥	٠,٥٠+	٢٨,٩٥	١٩٢٦
٤٠,٥	٦,٣٦+	٣٤,٨١	١٩٢٧
٠,٤	٠,٦٦+	٢٩,١١	١٩٢٨
١٠,٩	٣,٣٠-	٢٥,١٥	١٩٢٩
٦٥,٠	٨,٠٥+	٣٥,٥٠	١٩٣٠
٣٥٥,٩٢	٨٣,٧١	٨٥٣,٦٣	المجموع
٣٠	٣٠	٣٠	العدد
التباين = ١١,٨٦ الانحراف المعياري = ٣,٤٥	الانحراف المتوسط = ٢,٧	المتوسط = ٢٨,٤٥	

وبعد إجراء عمليات حسابية بسيطة باستخدام بيانات إنتاج خام الحديد في أربع دول من دول غرب أوروبا: بلجيكا، فرنسا، لوكسمبرج والمملكة المتحدة (جدول رقم ٤ - ١٢) يتم استخراج قيم كل من المتوسط الحسابي والانحراف المتوسط والانحراف المعياري ونسبة الانحراف المعياري إلى الانحراف المتوسط لإنتاج كل دولة كما هو مبين في الجدول رقم (٥ - ٥) التالي:

جدول (٥ - ٥)

الانحراف المتوسط والانحراف المعياري لإنتاج الحديد الخام في

بلجيكا، فرنسا، لوكسمبرج والمملكة المتحدة (١٩٣٨ - ١٩٥٧)

الدولة	المتوسط الحسابي (آلاف الأطنان)	الانحراف المتوسط (آلاف الأطنان)	الانحراف المعياري (آلاف الأطنان)	الانحراف المعياري المتوسط
بلجيكا	٣٤,٤٥	١٣,٥	١٥,٥٥	١,١٨
فرنسا	٩٣١٠,٢	٤٢٨٣,٢	٤٩٦٠,٠	١,١٢
لوكسمبرج	١٤٤٨,٨	٤٣٧,٣	٥٢٧,٢	١,٢٠
المملكة المتحدة	٤٤١٨,٩٥	٤٨٠,١	٦٥٦,٥	١,٣٧

ويلاحظ من الجدول السابق أن الانحراف المعياري، في كل الحالات، أكبر من الإنحراف المتوسط، كما أن النسبة بينهما تتراوح بين ١,١٦ و ١,٣٧. ويرجع السبب في ذلك إلى أنه أثناء عملية تربيع قيم المفردات ثم أخذ الجذر التربيعي لمجموع مربعات هذه القيم، فإن الانحرافات الكبيرة تحمل وزناً متزايداً، بينما تحمل الانحرافات الصغيرة وزناً قليلاً، ومعنى ذلك أن الانحراف المعياري يظل دائماً أكبر من الانحراف المتوسط. وتتوقف درجة الاختلاف بين الانحرافتين على التكرار النسبي للقيم ومدى الانحراف الذي يحدده أكبر وأصغر الانحرافات الفردية. وبالإضافة إلى ذلك فإن هذه الخاصية للانحراف المعياري تعكس حقيقة أن بيانات إنتاج الحديد في دول غرب أوروبا (جدول رقم ٤ - ١٢) تتصف بأنها غير متماثلة

فى توزيعها (ملتوية)، بينما تتصف بيانات كمية المطر السنوى فى مرصد Bidston - انجلترا بأنها تتوزع توزيعاً متماثلاً (أو معتدلاً) Normally distributed تقريباً.

طرق أخرى لحساب التباين والانحراف المعيارى:

سبق أن لاحظنا أن عملية إستخراج التباين والانحراف المعيارى لمجموعة من البيانات (معادلات رقم: ٥ - ٤، ٥ - ٥، ٤ - ٥، ٧ - ٥، ٨ - ٥) تكتنفها بعض الصعوبات الحسابية، ولذلك إذا كانت هناك عمليات أخرى لاختصار طرق حسابهما فقد يساعد ذلك على سهولة وتبسيط طرق استخراجهما من البيانات المتاحة. وهناك طريقتان يمكن بهما تبسيط عمليات استخراج التباين والانحراف المعيارى: الطريقة الأولى تعتمد على إجراء تعديل جبرى فى معادلة التباين وتعرف باسم طريقة الفرق بين المربعين، والطريقة الأخرى هدفها تقليل عدد الحسابات وخطواتها وتعرف باسم طريقة التربيع أو طريقة الماكينة. وسنشرح فيما يلى كل طريقة على حدة.

طريقة الفرق بين المربعين:

سبق أن ذكرنا أن التباين هو متوسط مربع انحرافات قيم المفردات عن متوسطها الحسابى أو ما يعرف «بالانحراف التربيعى المتوسط» ومعادلته هى:

$$\frac{\sum (s - \bar{s})^2}{n} = (ع٢)$$

ويظهر أن العنصر الرئيسى فى تلك المعادلة هو $(s - \bar{s})$ والذى يطلق عليه جبرياً «الفرق بين المربعين»، ويمكن أن يكتب ذلك تفصيلاً بالصورة التالية:

$$(s - \bar{s})^2 = (s - \bar{s})(s - \bar{s})$$

$$= s^2 - 2s\bar{s} + \bar{s}^2$$

ويقسمة مفكوك القوس على عدد المفردات (ن) يمكن إعادة كتابة معادلة التباين على النحو التالى:

$$\text{التباين (ع-٢)} = \frac{\text{مجم}^2 \text{س}}{\text{ن}} - \frac{\text{مجم}^2 \text{س}}{\text{ن}} + \frac{\text{مجم}^2 \text{س}}{\text{ن}} \dots (٥ - ١١)$$

وبالرغم من أنه يمكن إختصار كل جزء بمفرده من هذه الصورة المطولة لمعادلة التباين إلا أن عملية استخراج التباين بهذه المعادلة تبدو أكثر تعقيداً، ولذا يمكن تبسيط المكونات المفردة لهذه المعادلة عن طريق إستبدالها بمكونات أخرى.

فقد سبق القول أن $\bar{\text{س}}$ (المتوسط الحسابي) - $\frac{\text{مجم}^2 \text{س}}{\text{ن}}$ ، وعلى ذلك فإنه يمكن وضع $(\bar{\text{س}})$ بدلاً من $(\frac{\text{مجم}^2 \text{س}}{\text{ن}})$ وطالما أن قيمة $(\bar{\text{س}})$ ثابتة فإنها ستظل كذلك في بقية المعادلة، وبناء على ذلك إذا أضفنا قيمة $(\bar{\text{س}})$ بعدد (ن) من المرات ثم

قسمت بواسطة (ن) فإن الإجابة ستكون $\bar{\text{س}}$ ، أى أن :

$$\bar{\text{س}} = \frac{\text{مجم}^2 \text{س}}{\text{ن}}$$

ونحاول الآن تبسيط معادلة التباين (٥ - ١١) باستبدال $\bar{\text{س}}$ لكل من

$$\frac{\text{مجم}^2 \text{س}}{\text{ن}} \quad ، \quad \frac{\text{مجم}^2 \text{س}}{\text{ن}}$$

$$\text{ع-٢)} = \frac{\text{مجم}^2 \text{س}}{\text{ن}} - \bar{\text{س}}^2 \times \text{س} + \bar{\text{س}}^2$$

$$= \frac{\text{مجم}^2 \text{س}}{\text{ن}} - \bar{\text{س}}^2 + \bar{\text{س}}^2$$

$$= \frac{\text{مجم}^2 \text{س}}{\text{ن}} - \bar{\text{س}}^2 \dots \dots \dots (٥ - ١٣)$$

وبما أن الانحراف المعياري هو الجذر التربيعي للتباين فإن صيغة الانحراف المعياري تكتب كما يلي:

$$\text{الانحراف المعياري (ع-٣)} = \sqrt{\frac{\text{مجم}^2 \text{س}}{\text{ن}} - \bar{\text{س}}^2} \dots \dots (٥ - ١٣)$$

ويلاحظ من المعادلتين (٥ - ١٢، ٥ - ١٣) أنهما لا يتطلبان إلا حسابات قليلة، إذ يتم تربيع قيمة كل مفردة على حدة ثم تجمع مربعات هذه القيم وتقسم

على عدد المفردات ثم نطرح منها مربع المتوسط الحسابي للقيم فنحصل على التباين. وللحصول على الانحراف المعياري نستخرج الجذر التربيعي لقيمة التباين كما ذكرنا آنفاً.

طريقة الترييع (طريقة الماكينة):

يمكن أيضاً حساب مجموع مربع الانحرافات عن المتوسط الحسابي عند حساب التباين لمجموعة من البيانات باتباع الخطوات التالية:

مجموع مربع الانحرافات عن المتوسط الحسابي = $\sum (x_i - \bar{x})^2$ ، وهو المكون الرئيسي لمعادلة التباين (معادلة رقم ٥ - ٥) ويساوي:

$$\begin{aligned} \sum (x_i - \bar{x})^2 &= \sum (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \end{aligned}$$

وبالتعويض عن قيمة \bar{x} التي تساوي $\frac{\sum x_i}{n}$ في المعادلة السابقة:

$$\begin{aligned} \sum (x_i - \bar{x})^2 &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \\ &= \sum x_i^2 - 2\bar{x} \sum x_i + \sum \bar{x}^2 \end{aligned}$$

$$= \text{محد س}^2 - \frac{\text{محد س}^2}{n} \dots \dots \dots (15 - 5)$$

ويطلق على القيمة $\frac{\text{محد س}^2}{n}$ بعامل التصحيح، أى أن مجموع مربع الانحرافات عن المتوسط الحسابى يساوى مجموع مربع كل قيمة من قيم المفردات مطروحاً منه عامل التصحيح وهو عبارة عن مربع مجموع القيم مقسوماً على عددها. وعلى ذلك فإن صيغة كل من التباين والانحراف المعياري ستكون حيثئذ:

$$\text{التباين (ع2)} = \frac{1}{n} \left[\text{محد س}^2 - \frac{\text{محد س}^2}{n} \right] \dots (16 - 5)$$

$$\text{الانحراف المعياري} = \sqrt{\frac{1}{n} \left(\text{محد س}^2 - \frac{\text{محد س}^2}{n} \right)} \dots (17 - 5)$$

ويلاحظ أيضاً أن تطبيق أى من المعادلتين يتطلب حسابات أقل من طريقة الفرق بين المربعين السابق وشرحها، ذلك لأن كل قيمة مفردة يتم تربيعها وتجمع مربعات هذه القيم ثم تجمع قيم المفردات ويتم تربيع هذا المجموع ويقسم على عدد المفردات وي طرح الناتج من مجموع مربعات القيم ثم يقسم الناتج الأخير على عدد المفردات فنحصل على التباين، وبأخذ الجذر التربيعى له نحصل على الانحراف المعياري.

ولتوضيح تطبيق الطرق الثلاث السابقة لحساب التباين والانحراف المعياري نستعين بالمثال التالى:

مثال: فى دراسة لبيان مجال نفوذ مدينة ما، إتخذت خدمات السكك الحديدية بين تلك المدينة والمدن المجاورة معياراً لذلك. ولنفترض أن هناك ٢٥ مدينة تتم خدمتها بالقطارات، وأن عدد القطارات اليومية لكل من هذه المدن يختلف كما هو مبين فى العمود الثانى من الجدول رقم (٥ - ٦)، وبعملية حسابية بسيطة يمكن

إيجاد متوسط عدد القطارات فى اليوم الواحد بين المدينة، موضع الدراسة - وبين أى مدينة مجاورة، ووجد أنه ٩,٦ قطاراً فى اليوم. ومن الواضح أن هناك مدناً يزيد عدد القطارات التى تصل إليها عن هذا المتوسط زيادة كبيرة وأخرى يقل عدد القطارات عنه قلة واضحة، أى أنها بيانات تمثل متغيراً غير متصللاً Discrete Variable، ولذا فمن المفيد معرفة إنتشار (تشتت) قيم المفردات حول هذا المتوسط، وأدق المقاييس لمعرفة ذلك هو الانحراف المعيارى.

وبين الجدول رقم (٥-٦) طرق حساب الانحراف المعيارى عن طريق حساب التباين أولاً بالمعادلات التى سبق ذكرها من قبل وهى:

$$\text{التباين} = \frac{\text{مجم} (س - \bar{س})^2}{ن}$$

(الطريقة الأولى)

$$\text{التباين} = \frac{\text{مجم} س^2}{ن} - \bar{س}^2$$

(الطريقة الثانية)

$$\text{التباين} = \frac{1}{ن} \left[\frac{\text{مجم} (س^3)}{ن} - \bar{س}^3 \right]$$

(الطريقة الثالثة)

وكما هو موضح تعطى كل طريقة نفس قيمة التباين أى ٢٦,٠، وبذلك يكون الانحراف المعيارى فى كل ٥,١ قطاراً.

جدول رقم (٦٥)
حساب الانحراف المعياري لعدد القطارات اليومية بين مدينة
مركزية ومدن مجاورة باستخدام ثلاث طرق مختلفة

الطريقة الثانية والثالثة	الطريقة الأولى		عدد القطارات يومية (س)	المدن المجاورة
	الانحراف (س - س) ^٢	الانحراف (س - س)		
١	٧٣,٩٦	٨,٦-	١	١
٤	٥٧,٧٦	٧,٦-	٢	٢
٩	٤٣,٥٦	٦,٦-	٣	٣
٩	٤٣,٥٦	٦,٦-	٣	٤
١٦	٣١,٣٦	٥,٦-	٤	٥
٢٥	٢١,١٦	٤,٦-	٥	٦
٣٦	١٢,٩٦	٣,٦-	٦	٧
٣٦	١٢,٩٦	٣,٦-	٦	٨
٦٤	٢,٥٦	١,٦-	٨	٩
٦٤	٢,٥٦	١,٦-	٨	١٠
٦٤	٢,٥٦	١,٦-	٨	١١
١٠٠	٠,١٦	٠,٤+	١٠	١٢
١٠٠	٠,١٦	٠,٤+	١٠	١٣
١٠٠	٠,١٦	٠,٤+	١٠	١٤
١٠٠	٠,١٦	٠,٤+	١٠	١٥
١٢١	١,٩٦	١,٤+	١١	١٦
١٢١	١,٩٦	١,٤+	١١	١٧
١٤٤	٥,٧٦	٢,٤+	١٢	١٨
١٤٤	٥,٧٦	٢,٤+	١٢	١٩
١٩٦	١٩,٣٦	٤,٤+	١٤	٢٠
٢٢٥	٢٩,١٦	٥,٤+	١٥	٢١
٢٢٥	٢٩,١٦	٥,٤+	١٥	٢٢
٢٨٩	٥٤,٧٦	٧,٤+	١٧	٢٣
٣٦١	٨٨,٣٦	٩,٤+	١٩	٢٤
٤٠٠	١٠٨,١٦	١٠,٤+	٢٩	٢٥
٢٩٥٤	٦٥٠,٠٠		٢٤٠	المجموع

∴ المتوسط الحسابى (س) = $\frac{240}{20} = 12$
الطريقة الأولى :

$$26,0 = \frac{650}{20} = \text{التباين}$$

الطريقة الثانية :

$$\text{التباين} = \frac{2904}{20} - 2(12, 6)$$

$$26,0 = 118,16 - 92,16 =$$

الطريقة الثالثة :

$$\text{التباين} = \frac{1}{20} \left(\frac{2(240)}{20} - 2904 \right)$$

$$26,0 = 650 \times \frac{1}{20} =$$

والانحراف المعياري في كل الحالات $\sqrt{26} = 5,1$ قطاراً

ايجاد التباين والانحراف المعياري للبيانات المبوبة:

هناك عدة طرق لحساب التباين والانحراف المعياري للبيانات المبوبة في جداول التوزيعات التكرارية ولكنها لا تختلف عن طريقة حسابهما للبيانات غير المبوبة. ويحسب التباين بالصيغة التالية:

$$\text{التباين (ع-2)} = \frac{\text{محدك (س-س)}}{\text{محدك}} \dots \dots (18-4)$$

حيث محدك = مجموع التكرارات، م هي مراكز الفئات ولصعوبة استخدام هذه الصيغة يمكن وضعها على الصورة التالية:

$$(ع٢) = \frac{\text{مجم ك}^2}{\text{مجم ك}} - \left(\frac{\text{مجم ك}}{\text{مجم ك}} \right)^2 \dots \dots (١٩ - ٥)$$

ويكون الانحراف المعياري:

$$(ع٣) = \sqrt{\frac{\text{مجم ك}^2}{\text{مجم ك}} - \left(\frac{\text{مجم ك}}{\text{مجم ك}} \right)^2} \dots \dots (٢٠ - ٥)$$

وتسمى هذه الصيغة لحساب الانحراف المعياري بالطريقة المطولة.
ويبين الجدول رقم (٥ - ٧) خطوات حساب التباين والانحراف المعياري من جدول التوزيع التكراري لعدد القطارات اليومية بين مدينة مركزية ومدن مجاورة الذي تم إعداده على أساس تبويب البيانات الموضحة في الجدول رقم (٥-٦) في فئات كما هو متبع في إعداد الجداول التكرارية، وذلك مع إبقاء المدى المطلق Range للقيم في أى فئة صغيراً حتى يقل مدى الخطأ الناتج عن التعميم، ومع عدم تجاوز عدد الفئات عن عشرة فئات.

جدول رقم (٥ - ٧)

حساب التباين والانحراف المعياري من البيانات المبوبة لعدد القطارات اليومية بين مدينة مركزية ومدن مجاورة المين في الجدول رقم (٥ - ٦)

الفئات	الحدود الحقيقية للفئات	مراكز الفئات (م)	التكرار (ك)	م × ك	م ^٢ × ك
١ - ٢	٠,٥ - ٢,٥	١,٥	٢	٣,٠	٤,٥٠
٣ - ٤	٢,٥ - ٤,٥	٣,٥	٣	١٠,٥	٣٦,٧٥
٥ - ٦	٤,٥ - ٦,٥	٥,٥	٣	١٦,٥	٩٠,٧٥
٧ - ٨	٦,٥ - ٨,٥	٧,٥	٣	٢٢,٥	١٦٨,٧٥
٩ - ١٠	٨,٥ - ١٠,٥	٩,٥	٤	٣٨,٠	٣٦١,٠٠
١١ - ١٢	١٠,٥ - ١٢,٥	١١,٥	٤	٤٦,٠	٥٢٩,٠٠
١٣ - ١٤	١٢,٥ - ١٤,٥	١٣,٥	١	١٣,٥	١٨٢,٢٥
١٥ - ١٦	١٤,٥ - ١٦,٥	١٥,٥	٢	٣١,٠	٤٨٠,٥٠
١٧ - ١٨	١٦,٥ - ١٨,٥	١٧,٥	١	١٧,٥	٣٠٦,٢٥
١٩ - ٢١	١٨,٥ - ٢٠,٥	١٩,٥	٢	٣٩	٧٦٠,٥٠
			٢٥	٢٣٧,٥	٢٩٢٠,٢٥

ولحساب التباين والانحراف المعياري للبيانات السابقة نجد أن:

$$\text{التباين (ع-٢)} = \frac{٢٩٢٠,٢٥}{٢٥} - \left(\frac{٢٣٧,٥}{٢٥} \right)^٢$$

$$= ٩٠,٢٥ - ١١٦,٨١ = ٢٦,٥٦$$

والانحراف المعياري (ع-٣) $= \sqrt{٢٦,٥٦} = ٥,١٥٤$ قطاراً

وواضح أن القيمة التي حصلنا عليها للانحراف المعياري (٥,١٥٤) تختلف عن النتيجة السابقة (١,٥) من البيانات التفصيلية. وكما ذكرنا سابقاً عند حساب

المتوسط الحسابى من بيانات مبوبة، فإن الخطأ الناشئ عن العمليات الحسابية لإيجاد الانحراف المعياري للبيانات المبوبة يعتبر ضئيلاً على الرغم من العدد المحدود الفئات التوزيع الموضحة بالجدول، وهذا الخطأ من وجهة النظر الاحصائية يعد مقبولاً خصوصاً إذا عرفنا ما توفره طريقة تجميع البيانات وتبويبها من وقت فى إجراء العمليات الحسابية عنه مع الطريقة العادية للبيانات التفصيلية التى تتضمن عدداً كبيراً من قيم المفردات والتى غالباً ما تحتوى على قيم شاذة أو متطرفة.

ولتسهيل العمليات الحسابية لإيجاد التباين والانحراف المعياري من بيانات جدول توزيع تكرارى منتظم (أى الذى تتساوى فيه أطوال الفئات) نستخدم طريقة أخرى تعرف باسم طريقة الانحرافات المختصرة. وفى هذه الطريقة يختار وسط فرضى من بين مراكز الفئات ثم نحسب الانحرافات عن هذا الوسط الفرضى بعد إختصارها وذلك بقسمتها على طول الفئة ثم نربعها ونوجد متوسطها بعد ذلك، وهذا يعنى أننا طرحنا مقداراً ثابتاً من مراكز الفئات فيكون الانحراف المعياري للانحرافات (أى القيم بعد طرح الوسط الفرضى منها) هو نفسه الانحراف المعياري لمراكز الفئات الأصلية. وفى هذه الحالة يمكن حساب التباين والانحراف المعياري باستخدام الصيغة الآتية:

$$\text{التباين (ع-٢)} = L \times \left[\frac{\sum f_k^2}{\sum f_k} - \left(\frac{\sum f_k}{n} \right)^2 \right] \quad (٥ - ٢١)$$

$$\text{الانحراف المعياري (ع-٣)} = L \times \sqrt{\frac{\sum f_k^2}{\sum f_k} - \left(\frac{\sum f_k}{n} \right)^2} \quad (٥ - ٢٢)$$

حيث L = طول الفئة، \bar{C} = الانحراف المختصر عن الوسط الفرضى ويوضح الجدول رقم (٥ - ٨) خطوات إستخراج الانحراف المعياري من جدول التوزيع التكرارى لأطوال مائة رافد نهري (بالأمتار) المبينة بالجدول رقم (٢ - ١١).

جدول رقم (٨-٥)
حساب الانحراف المعياري لأطوال مائة رائدة نهري (منرى لأحد
الأخراض النهرية في إحدى المناطق باستخدام طريقة الانحرافات المختصرة

الانحرافات (ح)	التكرار ك	مراكز الفئات (س)	الحدود الحقيقية للفئات	الفئات (منرى)
١٩٠	١	٥٩٥	٦٤٥-٥٤٥	٦٤٥-٥٥٥
١٨	٢	٦٩٥	٧٤٥-٦٤٥	٧٤٥-٦٥٥
٣٦	٩	٧٩٥	٨٤٥-٧٤٥	٨٤٥-٧٥٥
٢٢	٢٢	٨٩٥	٩٤٥-٨٤٥	٩٤٥-٨٥٥
صفر	٣٣	٩٩٥	١٠٤٥-٩٤٥	١٠٤٥-٩٥٥
٨٨	٢٢	١٠٩٥	١١٤٥-١٠٤٥	١١٤٥-١٠٥٥
٧٢	١٦	١١٩٥	١٢٤٥-١١٤٥	١٢٤٥-١١٥٥
٣٢	٩	١٢٩٥	١٣٤٥-١٢٤٥	١٣٤٥-١٢٥٥
١٦	٤	١٣٩٥	١٤٤٥-١٣٤٥	١٤٤٥-١٣٥٥
١٨٠	٢	١٠٠		الجميع = منرى

ولإيجاد الانحراف المعياري باستخدام بيانات الجدول السابق نجد أن

$$\sqrt{\left[\left(\frac{2-}{100}\right) - \frac{180}{100}\right]} \times 100 = \text{عـ) الانحراف المعياري}$$

$$\sqrt{1,7996} \times 100 =$$

$$134,14 \text{ متراً.}$$

وتجدر الإشارة هناك إلى أنه يجب أن يسبق عملية استخراج الانحراف المعياري للبيانات المبوبة إجراء تحقيق وفحص صحة الحسابات في عملية الجدولة حتى نتفادى الوقوع في خطأ، وحتى يمكن عمل التصحيحات اللازمة قبل إيجاد قيمة هذا المقياس. وفي هذا الصدد يمكن تطبيق طريقة «اختبار تشارليير Charlier's Test» لمراجعة حساب الانحراف المعياري باستخدام المتطابقات:

$$\text{مجم ك} (ح + 1) = \text{مجم ك} ح + 2 + \text{مجم ك} ح + \text{مجم ك}$$

$$100 + (2-) 2 + 180 = 276$$

$$276 = 276$$

ويلاحظ أن طرفي المعادلة متساويان مما يدل على أن العمليات الحسابية في الجدولة قد تمت بدقة، وأن أى خطأ صغير يبقى بين الأرقام يمكن اعتباره خطأ ناتج عن التعميم والناتج بدوره عن اتساع الفئات التكرارية.

تصحيح التباين والانحراف المعياري:

عند حساب كل من التباين والانحراف المعياري للبيانات المبوبة فإنه يكون معرضاً لبعض الخطأ الناتج عن تجميع البيانات في فئات (أخطاء التجميع)، بمعنى أننا نفترض أن مفردات كل فئة تتجمع أو تمثلها قيمة واحدة هي مركز الفئة وهو افتراض تقريبي الغرض منه تسهيل العمليات الحسابية. وبما أن القيم الحقيقية للمفردات تختلف داخل الفئة الواحدة فإن الفرق بين القيم الحقيقية والقيم

المفروضة (مراكز الفئات) تؤدي إلى بعض الأخطاء في نتائج العمليات الحسابية، وبالتالي فإن نتائج المقاييس الإحصائية، ومن بينها مقاييس الانحراف المحسوبة من جدول توزيع تكرارى لا تنطبق تماماً على النتائج للمقاييس المناظرة المحسوبة من البيانات التفصيلية (راجع نتائج حساب كل من التباين والانحراف المعيارى لبيانات الجدولين رقم (٥ - ٦) (٥ - ٧). ولتعديل هذا النوع من الخطأ نستخدم الصيغة التالية:

التباين المعدل (المصحح) = التباين المحسوب من البيانات المبوبة - $\frac{\sum l^2}{12}$
حيث ل هي طول الفئة، ومعامل التصحيح $(\frac{\sum l^2}{12})$ المطروح يسمى تصحيح شبرد Sheppard's Correction، وبأخذ الجذر التربيعى لباقي الطرح من الصيغة السابقة ينتج الانحراف المعيارى المصحح. ففي المثال الموضح فى الجدول رقم (٥ - ٧) وجدنا أن التباين قيمته ٢٦,٥٦ وهو يختلف عن التباين للبيانات التفصيلية (قبل التبويب) وقيمته ٢٦,٠ وعلى هذا يكون التباين المعدل هو:

$$\frac{\sum l^2}{12} - 26,56 = \text{التباين المعدل}$$

$$26,23 = 0,33 - 26,56 =$$

ويكون الانحراف المعيارى المصحح $= \sqrt{26,23} = 5,13$ قطاراً.

ومن الواضح أن تصحيح شبرد لا يستخدم إلا فى حالة توزيعات المتغيرات المتصلة (المتغيرات المستمرة Continuous Variables) حيث «أطراف التوزيع» فيها تؤول تدريجياً إلى الصفر فى كلا الاتجاهين، أو بمعنى آخر لايسرى على التوزيعات النونية (على شكل حرف U) أو التوزيعات الشديدة الالتواء. ويختلف الإحصائيون فى متى وما إذا كان تصحيح شبرد يجب تطبيقه لتعديل التباين والانحراف المعيارى المحسوبين من بيانات مبوبة. إلا أنه يجب عدم تطبيقه إلا بعد فحص دقيق للبيانات إذ أنه كثيراً ما يؤدي إلى مبالغة فى التصحيح وهذا يؤدي بدوره إلى استبدال الخطأ

القديم بخطأ جديد. وعلى العموم فإنه يفضل عدم إجراء هذا التصحيح للجداول التكرارية التى تقل فيها عدد التكرارات عن ١٠٠٠ أو تقل بها عدد الفئات عن ٢٠ فئة.

مؤشرات الاختلاف (التباين) Variability Indices

ذكرنا أنه لمقارنة مجموعتين من البيانات يجب استخدام متوسطيهما أولاً ثم تشتيتهما بعد ذلك بواسطة استخدام عدة مقاييس احصائية للمتوسطات والتشتت. وكما سبق أن لاحظنا على هذه المقاييس أنها تعطى فى النهاية قيماً مطلقة (غير نسبية) من نفس وحدات الصفة المتغيرة المستخرجة منها (مثل عدد الأمتار، البوصات الأطنان، عدد السكان، عدد الدرجات: مئوية أو فهرنيهية بالنسبة لدرجة الحرارة مثلاً ... إلخ). وبما أنه من الطبيعى أن تختلف وحدات الصفة المتغيرة من مجموعة من القيم إلى مجموعة أخرى مما يؤدى بدوره إلى اختلاف وحدات مقاييس المتوسطات والتشتت، فإن المقارنة لاتصلح إلا بين المجموعات ذات النوع الواحد من الصفة المتغيرة للبيانات، فلا يعقل مثلاً أن نقارن بين كمية البضائع (بالأطنان) بأطوال السكك الحديدية (بالكيلو مترات) المنقولة عليها. وللتغلب على الاختلاف فى وحدات هذه المقاييس وحتى تكون المقارنة صحيحة ينبغي استخدام وحدات من نفس النوع أو استخدام أعداد مجردة خالية من الوحدات التى يمكن الحصول عليها بقسمة عددين من نوع واحد (أى من نفس الوحدات) أحدهما على الآخر. فمثلاً نقسم أحد مقاييس الانحراف على أقرب مقاييس المتوسطات إليه فى طبيعته ويعبر عن ناتج القسمة كنسبة مئوية (أى نضرب الناتج فى ١٠٠٪)، وبذلك نكون قد حولنا قيم مقاييس الانحراف من قيم مطلقة إلى قيم مئوية (بالنسبة للمتوسطات)، وهذه القيم المئوية هى التى نطلق عليها اسم «مؤشرات الاختلاف أو التباين». ونظراً للأهمية والفائدة الكبيرة لهذه المؤشرات عند عقد المقارنات بين المجموعات من البيانات فقد أصبحت تستخدم اليوم على نطاق كبير فى التحليل الكمي للجغرافية.

ومؤشرات الاختلاف أنواع منها مؤشر الاختلاف الذى يستخدم قيمة الوسيط والانحراف الربيعى وذلك على النحو التالى:

$$\frac{\text{الانحراف الربيعى}}{\text{الوسيط}} \times 100\% \quad \dots \quad (23 - 5)$$

والمؤشر الثانى للاختلاف هو ما يطلق عليه «مؤشر الاختلاف النسبى» Relative Variability Index ويستخدم قيمة المتوسط الحسابى والانحراف المتوسط ويمكن الحصول عليه من الصيغة الآتية:

$$\frac{\text{الانحراف المتوسط}}{\text{المتوسط الحسابى}} \times 100\% \quad \dots \quad (24 - 5)$$

أما إذا استخدمنا الانحراف المعيارى والمتوسط الحسابى فإننا نحصل على مؤشر للاختلاف يعرف باسم معامل الاختلاف Coefficient of Variation ويستخرج كالآتى:

$$\frac{\text{الانحراف المعيارى}}{\text{المتوسط الحسابى}} \times 100\% \quad \dots \quad (25 - 5)$$

وبعد معامل الاختلاف أكثر مؤشرات الاختلاف انتشاراً، ويستخدم فى الحالات التالية:

١- تؤخذ قيمة معامل الاختلاف كمؤشر أو دليل لمدى دقة القياس أو فى عملية الحصول على البيانات إذا أنه لايمكن أن نقارن بين الانحرافات المعيارية مباشرة لاختلاف قيمة المتوسطات الحسابية لهذه العمليات. ولكن يمكن القول أنه كلما قلت قيمة معامل الاختلاف كلما زادت دقة القياس والعكس صحيح. ونظراً لأن معامل الاختلاف للقياسات العقلية يختلف عن معامل الاختلاف

للقياسات المعملية فقد وضعت حدوداً لقيمة المعامل يجب أن لاتتعداها القياسات فى كل حالة وإلا كانت النتائج المستحصلة غير دقيقة. وتتراوح قيمة معامل الاختلاف للقياسات الحقلية ما بين ١٠ - ٢٠ ٪ بينما تقل للقياسات المعملية عن ١٠ ٪. وفى حالة القياسات الحقلية إذا قلت قيمة معامل الاختلاف للبيانات عن ١٠ ٪ فإنه يجب مراجعة قيم البيانات خوفاً من وقوع بعض الأخطاء فى العمليات الحسابية، أما إذا زادت قيمته عن ٢٠ ٪ فإنه يستدل من ذلك على أن عملية القياس قد تمت تحت ظروف غير ملائمة مما يترتب عليه استخلاص نتائج لايعول عليها كثيراً، ولذا يجب التأكد من شروط عملية القياس قبل إجرائها.

٢- يستخدم معامل الاختلاف لمقارنة الاختلافات بين توزيع عينتين لنفس الصفة المتغيرة ولكنها مقاسة بوحداث مختلفة، فمثلاً إذا كانت لدينا بيانات عن توزيع درجة الحرارة فى إحدى المناطق لفترتين مختلفتين وكان متوسط الحرارة للفترة الأولى هو ٢٣ درجة مئوية والانحراف المعيارى ٢ م، وكان متوسط درجة الحرارة للفترة الثانية هو ٨٦ درجة فهرنهايتية والانحراف المعيارى ٦ ف. ففى هذه الحالة لايمكننا أن نقارن بين الانحراف المعيارى للفترتين وذلك لاختلاف وحدات القياس ولكن يمكن التغلب على ذلك بمقارنة معاملى الاختلاف للفترتين.

$$\text{فمعامل الاختلاف للفترة الأولى} = \frac{2}{23} \times 100 = 8.7\%$$

$$\text{ومعامل الاختلاف للفترة الثانية} = \frac{6}{86} \times 100 = 6.97\%$$

أى أن الاختلاف فى درجة الحرارة للفترة الأولى أكبر منه للفترة الثانية.

٣- يستخدم معامل الانحراف أيضاً فى حالة مقارنة التوزيعات ذات الفروق الكبيرة

ومؤشرات الاختلاف أنواع منها مؤشر الاختلاف الذى يستخدم قيمة الوسيط والانحراف الربيعى وذلك على النحو التالى:

$$\frac{\text{الانحراف الربيعى}}{\text{الوسيط}} \times 100\% \quad \dots \quad (23 - 5)$$

والمؤشر الثانى للاختلاف هو ما يطلق عليه «مؤشر الاختلاف النسبى» Relative Variability Index ويستخدم قيمة المتوسط الحسابى والانحراف المتوسط ويمكن الحصول عليه من الصيغة الآتية:

$$\frac{\text{الانحراف المتوسط}}{\text{المتوسط الحسابى}} \times 100\% \quad \dots \quad (24 - 5)$$

أما إذا استخدمنا الانحراف المعيارى والمتوسط الحسابى فإننا نحصل على مؤشر للاختلاف يعرف باسم معامل الاختلاف Coefficient of Variation ويستخرج كالتالى:

$$\frac{\text{الانحراف المعيارى}}{\text{المتوسط الحسابى}} \times 100\% \quad \dots \quad (25 - 5)$$

ويعد معامل الاختلاف أكثر مؤشرات الاختلاف انتشاراً، ويستخدم فى الحالات التالية:

١- تؤخذ قيمة معامل الاختلاف كمؤشر أو دليل لمدى دقة القياس أو فى عملية الحصول على البيانات إذا أنه لا يمكن أن نقارن بين الانحرافات المعيارية مباشرة لاختلاف قيمة المتوسطات الحسابية لهذه العمليات. ولكن يمكن القول أنه كلما قلت قيمة معامل الاختلاف كلما زادت دقة القياس والعكس صحيح. ونظراً لأن معامل الاختلاف للقياسات العقلية يختلف عن معامل الاختلاف

للقياسات العملية فقد وضعت حدوداً لقيمة المعامل يجب أن لاتتعداها القياسات في كل حالة وإلا كانت النتائج المستحصلة غير دقيقة. وتتراوح قيمة معامل الاختلاف للقياسات الحقلية ما بين ١٠ - ٢٠ ٪ بينما تقل للقياسات العملية عن ١٠ ٪. وفي حالة القياسات الحقلية إذا قلت قيمة معامل الاختلاف للبيانات عن ١٠ ٪ فإنه يجب مراجعة قيم البيانات خوفاً من وقوع بعض الأخطاء في العمليات الحسابية، أما إذا زادت قيمته عن ٢٠ ٪ فإنه يستدل من ذلك على أن عملية القياس قد تمت تحت ظروف غير ملائمة مما يترتب عليه استخلاص نتائج لايعول عليها كثيراً، ولذا يجب التأكد من شروط عملية القياس قبل إجرائها.

٢- يستخدم معامل الاختلاف لمقارنة الاختلافات بين توزيع عيتين لنفس الصفة المتغيرة ولكنها مقاسة بوحدات مختلفة، فمثلاً إذا كانت لدينا بيانات عن توزيع درجة الحرارة في إحدى المناطق لفترتين مختلفتين وكان متوسط الحرارة للفترة الأولى هو ٢٣ درجة مئوية والانحراف المعياري ٢ م، وكان متوسط درجة الحرارة للفترة الثانية هو ٨٦ درجة فهرنهايت والانحراف المعياري ٦ ف. ففي هذه الحالة لايمكننا أن نقارن بين الانحراف المعياري للفترتين وذلك لاختلاف وحدات القياس ولكن يمكن التغلب على ذلك بمقارنة معامل الاختلاف للفترتين.

$$\text{فمعامل الاختلاف للفترة الأولى} = \frac{2}{23} \times 100 = 8,7\%$$

$$\text{ومعامل الاختلاف للفترة الثانية} = \frac{6}{86} \times 100 = 6,97\%$$

أى أن الاختلاف في درجة الحرارة للفترة الأولى أكبر منه للفترة الثانية.

٣- يستخدم معامل الانحراف أيضاً في حالة مقارنة التوزيعات ذات الفروق الكبيرة

بين متوسطاتها حتى ولو كانت بياناتها مقيسة بنفس وحدات القياس. مثال ذلك إذا كان المتوسط الحسابي والانحراف المعياري للملكية الزراعية في منطقتين هو: في المنطقة الأولى المتوسط = ١٠ أفدنة والانحراف المعياري ١,٥ فدناً، وفي المنطقة الثانية متوسط الملكية = ٥ أفدنة والانحراف المعياري ١,٠ فدناً. ولأول وهلة يمكن أن نقول، اعتماداً على القيم المطلقة للتشتت، أن الملكية في المنطقة الأولى أكثر اختلافاً عن المنطقة الثانية وذلك إذا أهملنا المتوسط الحسابي لكلا المنطقتين، ولكن ذلك لا يكون صحيحاً نظراً لاختلاف المتوسط الحسابي في كل منهما. أما إذا حسبنا الاختلافات بالنسبة للمتوسط الحسابي نجد أن:

$$\text{معامل الاختلاف للمنطقة الأولى} = \frac{1,5}{10} \times 100 = 15\%$$

$$\text{ومعامل الاختلاف للمنطقة الثانية} = \frac{1}{5} \times 100 = 20\%$$

وتشير النتيجة السابقة للاختلاف النسبي مقيساً بمعامل الاختلاف للملكية في المنطقتين أن المنطقة الثانية أكثر اختلافاً من المنطقة الأولى لو أدخلنا المتوسط الحسابي في الحساب عند تقييم النتائج، وهذه النتيجة تخالف تماماً النتيجة التي نحصل عليها بالاعتماد على القيم المطلقة لاختلاف أو تشتت الملكية في المنطقتين، أو بمعنى آخر بالاعتماد على الانحراف المعياري فقط لمقارنة درجة الاختلاف بينهما.

وكمثال تطبيقي، تم إجراء العمليات الحسابية الخاصة بكل مؤشر من مؤشرات الاختلاف السابق ذكرها باستخدام بيانات إنتاج خام الحديد في أربع دول من دول غرب أوروبا (جدول رقم ٤ - ١٢) واستخرجت قيم هذه المؤشرات ووضعت بالجدول التالي:

جدول رقم (٥ - ٩)
مؤشرات الاختلاف لإنتاج خام الحديد في بلجيكا وفرنسا ولوكسمبورج
والمملكة المتحدة في الفترة من ٣٨ - ١٩٥٧ (نسب مئوية)

الدولة	الانحراف المعياري (ألف طن)	مؤشر الاختلاف، الانحراف الربيعي المتوسط ١٠٠ ×	الاختلاف النسبي، الانحراف المتوسط المتوسط ١٠٠ ×	معامل الاختلاف، الانحراف المعياري المتوسط ١٠٠ ×
بلجيكا	١٥,٥٥	٤١,٣٠	٣٨,٢	٤٥,١
فرنسا	٤٩٦٠,٠	٤٤,٨٥	٤٦,٠	٥٣,٣
لوكسمبورج	٥٢٧,٢	٢٨,١٠	٣٠,٢	٣٦,٤
المملكة المتحدة	٦٥٦,٥	٦,٦٥	١٠,٨	١٤,٨٥

من الجدول السابق نستطيع استخلاص عدة نقاط هامة توضح العلاقة بين مؤشرات الاختلاف الثلاثة وبينها وبين الانحراف المعياري وذلك على النحو التالي:

أولاً: أن ترتيب الدول الأربع حسب درجة الاختلاف لا يختلف من مؤشر لآخر، فالترتيب في كل حالة، باستثناء الانحراف المعياري، هو: فرنسا، بلجيكا، لوكسمبورج، والمملكة المتحدة.

ثانياً: أن قيمة درجة الاختلاف تختلف اختلافاً واضحاً بين طرق المؤشرات الثلاثة. ويعني ذلك أنه يجب عند توضيح درجة الاختلاف بين مجموعات البيانات أن نذكر الطريقة التي تم بواسطتها حساب قيمة الاختلاف (التباين) بينها. ويمكن القول أن الطريقة التي تستخدم كل من الانحراف الربيعي والوسيط تعطى قيمة أقل من الطريقتين الأخرتين، وذلك عند إهمال قيمة مؤشر الاختلاف لدولة بلجيكا التي تؤثر على الترتيب السابق لدرجة الاختلاف. ويرجع السبب في ذلك إلى أن عيوب كل من الوسيط والانحراف الربيعي

التي ذكرناها سابقاً يتضمنها أيضاً مؤشراً لاختلاف الذى يعتمد عليهما فى حسابه. ونستطيع من الجدول إيجاد علاقة اعتبارية تربط بين معامل الاختلاف ومؤشر الاختلاف النسبى، إذ يلاحظ أن معامل الاختلاف أكبر من مؤشر الاختلاف النسبى وأن الفرق بين قيم كل منهما قد يصل إلى ٢٥ ٪. وهى بذلك تشبه العلاقة التى أوضحناها سابقاً بين الانحراف المتوسط والانحراف المعيارى.

ثالثاً: أن ترتيب الدول الأربع على أساس الانحراف المعيارى فقط يختلف عن الترتيب السابق مع مؤشرات الاختلاف الثلاثة. فمثلاً على الرغم من أن الانحراف المعيارى للمملكة المتحدة يعد الثانى من حيث الترتيب بينما نجد أن قيمة معامل الاختلاف لها تعد الأقل بين قيم معامل الاختلاف للدول الثلاث الأخرى. وإن دل هذا على شئ فإنما يدل على أن إهمال إدخال المتوسط الحسابى فى الحساب يؤثر تأثيراً كبيراً فى تقييم النتائج، أو بمعنى آخر إذا أردنا تقييم الاختلاف بالنسبة للمتوسط الحسابى فإن المقارنة تكون أكثر وضوحاً من مقارنة الانحرافات المعيارية نفسها.

القيم (الوحدات) المعيارية (Standard Values (Units)

سبق أن ذكرنا أن معامل الاختلاف يستخدم كمقياس إحصائى للمقارنة بين تشتى مجموعتين من البيانات العينية. أما إذا أردنا المقارنة بين قيمتين فى مجموعتين من البيانات العينية. أما إذا أردنا المقارنة بين قيمتين فى مجموعتين مختلفتين فإننا يجب أن نقارن بين موضع كل من هاتين القيمتين فى التوزيع الخاص بهما عن طريق إيجاد بعدهما عن المتوسط الحسابى بدلالة وحدات من الانحراف المعيارى. ويتم ذلك بقسمة انحراف كل قيمة عن متوسطها الحسابى على انحرافها المعيارى، ويمكن أن يكتب ذلك بالصيغة التالية:

$$\frac{s - \bar{s}}{e} = \text{القيمة المعيارية}$$

حيث s هي قيمة المتغير، s هي متوسط قيم المتغير، e هي الانحراف المعياري لقيم المتغير.

وتستخدم القيم المعيارية أيضاً في المقارنة بين المتغيرات التي تقاس على مقاييس مزدوجة، كأن تكون قياسات المسافات بالكيلو مترات والأميال، أو تكون قياسات درجة الحرارة بالدرجات المئوية والفهرنهايتية، أو تكون قياسات المطر بالسنتيمترات والبروصات، وذلك للوقوف على معرفة توزيع وتشتت قيم المتغير في كل من حالات القياس المختلفة. فمثلاً إذا كانت لدينا المتوسطات الشهرية لدرجة الحرارة في المراصد الجوية في اقليم ما خلال شهور فصل الصيف (يونيو - أغسطس) وأردنا معرفة أى الشهور أنسب في درجة حرارته في هذا المرصد بالنسبة لبقية مواقع المراصد الأخرى، فإنه يمكن حساب القيمة المعيارية لكل شهر من الشهور ونقارنها مع بعضها البعض وذلك على النحو التالي:

يوليو	يوليو	أغسطس
المعدل الشهري في مراصد الاقليم	٢٥,٤	٢٦,٦
درجة الحرارة الشهرية في المرصد	٢٤,١	٢٥,٧
الانحراف المعياري	١	١,٥

وتكون القيمة المعيارية لكل شهر على حدة هي:

$$1,3 = \frac{1,3}{1} = \frac{25,4 - 24,1}{1} = \text{القيمة المعيارية لشهر يونيو}$$

$$0,66 = \frac{1}{1,5} = \frac{26,7 - 25,7}{1} = \text{القيمة المعيارية لشهر يوليو}$$

$$0,1 = \frac{0,2}{2} = \frac{26,6 - 26,4}{2} = \text{القيمة المعيارية لشهر أغسطس}$$

وبما أن القيمة المعيارية لشهر أغسطس (٠, ١-) هي أعلى القيم فإن درجة حرارة المرصد قيد البحث في هذا الشهر كانت أفضل أو أكثر تحملاً من حرارة الشهرين الآخرين بالنسبة لبقية المراصد في الاقليم، ثم يأتي بعد ذلك شهر يوليو فشهر يونيو من حيث هذه الأفضلية.

والخلاصة أن مقاييس الانحراف ومؤشرات الاختلاف السابقة تعد ذات فائدة كبيرة في التحليلات الاحصائية الجغرافية - أى في المقارنات المختلفة بين المناطق والأقاليم - وذلك عن طريق تمثيل هذه المؤشرات على خرائط أو في شكل رسوم بيانية. وما تجدر الإشارة إليه هنا أنه لا يوصى باستخدام كل من الانحراف الربيعي والانحراف المتوسط وما يرتبط بهما من مؤشرات للاختلاف (لتباين) بدرجة كبيرة في التحليل الكمي للبحوث الجغرافية إذ يعانى كل منهما من نفس عيوب مقاييس المتوسطات التي تدخل في تحديدها. ولكن يمكن القول بأن معظم الطرق والأساليب الكمية التي سيرد ذكرها في متن الفصول القادمة تعتمد اعتماداً يكاد يكون كلياً على مقياس المتوسط الحسابي ومقاييس التباين (الانحراف المعياري، ومعامل الاختلاف) والقيم المعيارية وهى المقاييس التي تستخدم، كما ذكرنا، كمؤشرات إحصائية لبيان طبيعة البيانات وكيفية توزيع مفرداتها، والتي سيبرز جوهرها ومضمونها بصورة واضحة عند استخدامها، أيضاً، في التحليل الكمي للأمثلة الجغرافية فيما بعد.

الفصل السادس
مؤشرات التركيز
(الالتواء والتفرطح)

الفصل السادس.

مؤشرات التركيز

(الالتواء والتفرطح)

سبق أن أوضحنا فى الفصلين السابقين كيفية حساب مقاييس النزعة المركزية (أو المتوسطات) وكذلك حساب مقاييس التشتت وفائدتها فى وصف التوزيعات المختلفة وإعطاء فكرة عن حجم البيانات الاحصائية وتشتتها حول متوسطها الحسابى. ولكن هذه المقاييس لا تكفى فى وصف التوزيعات ومقارنتها بعضها البعض، إذ أنه قد يتساوى توزيعان تكراريان من حيث المتوسط الحسابى والانحراف المعياري ولكنهما يختلفان من حيث بعد المنحنى التكرارى للتوزيع عن التماثل. أو بمعنى آخر أن هذه المقاييس لا ينتج عنها معلومات عن خصائص ومميزات شكل Shape التوزيع التكرارى، بل تقيس درجة إتساع «عرض Width» المنحنى التكرارى للتوزيع. ويمكن تحديد بعد أو اقتراب التوزيع من التماثل من خلال شكل المنحنى التكرارى للتوزيع ومقارنته بمنحنى متماثل. كما يمكن كذلك تحديد شكل المنحنيات الوحيدة القمة من حيث تفرطحها أو درجة تديبها. فقد تتساوى بعض المنحنيات المتشابهة فى وجود قمة واحدة لها فى بعض الخصائص التى يمكن الحصول عليها بمقاييس النزعة المركزية والتشتت أو حتى بمقاييس عدم التماثل أو الالتواء إلا أنها تختلف فى شكل قمته. لكل ذلك فإننا سنلقى الضوء فى هذا الفصل على مؤشرين من المؤشرات الإحصائية التى تقيس اتجاهات تركيز القيم هى: الالتواء والتفرطح لما لهما من أهمية لا تقل عن أهمية تحديد المتوسط.

تحديد المتوسط الحسابي والتشتت في تشخيص التوزيعات وتحديد خصائصها وملاحظاتها.

الإلتواء Skewness

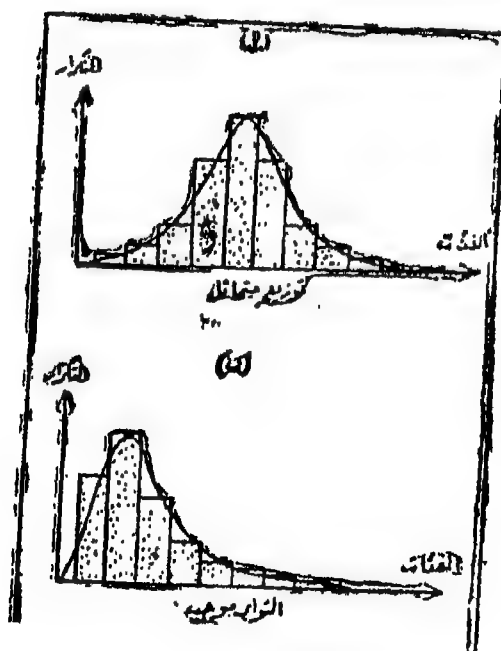
يعرف الإلتواء بأنه عبارة عن بعد لامدرج التكرارى أو المنحنى التكرارى عن التماثل حول المتوسط الحسابى للتوزيع، وهو بذلك يقيس اتجاه تركيز القيم كما يحدد مناطق وجود بعض القيم المتطرفة فى التوزيع التكرارى. وباختصار فإنه يعطينا أسلوباً لتوزيع القيم وتحديد الامتداد الذى يتركز فيه القدر الأكبر منها على أحد جانبي المتوسط الحسابى ومقارنته بتوزيع القيم فى التوزيعات التكرارية المتماثلة. ويجدر الإشارة إلى أنه توجد بعض الظواهر الجغرافية التى تتوزع بياناتها توزيعاً متماثلاً أو قريباً جداً من التماثل حيث تتركز معظم القيم عند منتصف التوزيع، ولكن لا ينطبق ذلك على الكثير من ظواهر الجغرافية الطبيعية التى تشير طبيعياً بأثر فعل عامل طبيعى ينتج عنه بيانات تتوزع أو تتركز فى أحد أطراف التوزيع عنه فى الطرف الآخر مما يبعدها عن التماثل، وكما سبق ذكره أنه عندما لخطنا البيانات الاحصائية على هيئة جداول توزيعات تكرارية ورسمناها المدرجات التكرارية (الهستوجرام)، فإن الأخيرة قد تأخذ الأشكال الثلاثة الآتية:

١- قد تتزايد التكرارات تدريجياً إلى أعلى حتى تصل إلى أكبر قيمة لها ثم تناقص التكرارات إلى أسفل وينسب تكاد تكون متساوية وتظهر هذه الخاصية بوضوح إذا أكثرنا من عدد الفئات وعدد المفردات وصغرنا من طول الفئة، وينعكس ذلك على شكل التوزيع الناتج حيث نحصل على مدرج تكرارى يسمى «مدرج تكرارى متماثل» (شكل رقم ٦-١) يوحى بأن المنحنى التكرارى المماثل له منحنى متماثل، إذ أن قمة المنحنى فى منتصف التوزيع تماماً، وأن طرفاه ينطبقان إلى درجة كبيرة على بعضهما عند المحور. ويعنى هذا أنه لا توجد قيم متطرفة أو شاذة سواء كانت كبيرة أو صغيرة تسبب ابتعاد أحد الأطراف عن الآخر، أو تؤدي إلى التواء التوزيع وعدم تحقيق تطابق طرفيه.

٢- قد يبدأ المدرج التكرارى بتركز كبير للتكرارات فى إطار الفئات الأولى

(الصغيرة) للتوزيع ثم يقل تركزها ويتناقص تدريجياً في إطار الفئات الأخيرة (الكبيرة) وبصورة متطرفة مما يسبب عدم تطابق طرفي التوزيع، ويصبح شكل المدرج كما في الشكل رقم (٦-١ ب). ويقال إحصائياً أن المدرج التكرارى للتوزيع ملتو لتواء موجباً أو ملتو ناحية اليمين حيث أن ذيل المنحنى الذى يمثل هذا التوزيع التكرارى يتجه ناحية اليمين من تأثير تطرف القيم فى الفئات الكبيرة.

٣- قد تبدأ التكرارا فى الفئات الأولى من التوزع صغيرة، مثل هذه التكرارات يمكن اعتبارها أيضاً متطرفة أو شاذة، ثم تزداد هذه التكرارات فجأة فى إطار الفئات الكبيرة. فى مثل هذه الحالة يقال أن التوزيع ملتو ناحية اليسار، وأن المنحنى الذى يمثل هذا التوزيع منحنى ملتو سالب حيث أن ذيل المنحنى أو الطرف الأيسر للتوزيع يتجه نحو اليسار أبعد من إتجاه الطرف الأيمن (شكل رقم ٦-١ ح).



شكل رقم (٦-١)

أشكال المدرجات التكرارية وأنواع التوزيعات التكرارية التى تمثلها

والإلتواء بهذا المفهوم السابق يعبر عن عدم التماثل في التوزيعات التكرارية، ولذا فإن تحديد مقدار ونوع الإلتواء يعد غابة في الأهمية خصوصاً إذا علنا أنه قد يكون هناك بعض التوزيعات التي تتساوى في متوسطاتها الحسابية وأيضاً في تشتتها في الوقت الذي تختلف فيه تماماً في التوائها. ومن هنا تميز التوزيعات عن بعضها، فقد يكون التواء التوزيعات في اتجاه واحد ولكنه يختلف في مقداره، أو قد تكون درجة الإلتواء في التوزيعات متساوية ولكنها تختلف في مقدار هذا الإلتواء، أو قد تكون درجة التوائها متساوية ولكنها تختلف في النوع، ويمكن تحديد درجة الإلتواء (بسيط - متوسط - حاد) وأيضاً نوع الإلتواء (موجب - سالب) من خلال شكل المدرج أو المنحنى التكرارى للتوزيع ومقارنته بمدرج أو بمنحنى متماثل. إلا أن هذا الأسلوب لا يعطى قياساً دقيقاً لتحقيق هذا الغرض، ولذا فمن المستحسن استخدام بعض المقاييس الكمية التي تقيس الإلتواء.

مقاييس الإلتواء:

لما كان التوزيع التكرارى المتماثل يتميز بانطباق كل من الوسط الحسابى والوسط والمنوال بعضها على بعض، فإن وجود فرق بين هذه المقاييس إنما ينتج عنه التواء فى المنحنى. وبناء على ذلك فإنه يمكن استخدام الفروق بين قيم هذه المتوسطات لقياس للإلتواء. ونظراً لأن الفروق بين المتوسطات الثلاثة تتخذ شكل الوحدات المعيارية (القياسية) التي تختلف من توزيع إلى آخر فإن هذه الفروق لا تقيس الإلتواء تماماً لأنه قد تكون الفروق كبيرة، والإلتواء بسيطاً، لزيادة تشتت البيانات، وقد تكون الفروق صغيرة والإلتواء حاداً لصغر تشتت لبيانات. ولذلك فأننا يجب أن ننسب الفروق بين المتوسط الحسابى والوسيط أو المتوسط الحسابى والمنوال إلى أحد مقاييس التشتت أو من نفس نوع مقاييس النزعة المركزية ونطلق على المقياس الناتج إسم «معامل الإلتواء». وتجدر الإشارة إلى أن معامل الإلتواء يجب أن يحقق شرطين أساسيين هما:

- ١- أن يكون هذا المعامل مساوياً للصفر وذلك بالنسبة للتوزيعات المتماثلة.
- ٢- أن لا تكون قيمة المعامل ذات تمييز معين، أو لا تتوقف على الوحدات التي

تقاس بها قيم المتغير. أو بعبارة أخرى أن تكون قيمته عدداً بحتاً. وعليه يمكن استعراض المقاييس الشائعة لحساب معامل الالتواء فيمايلي:

$$(١) \text{ معامل الالتواء} = \frac{\text{الوسط الحسابي} - \text{المنوال}}{\text{الانحراف المعياري}} \text{ أو}$$

$$\frac{\bar{S} - \bar{O}}{E} = \text{م.ت.١}$$

وفي التوزيعات الملتوية ناحية اليمين يقع المتوسط على نفس جانب المنوال أو في اتجاه القيم الكبيرة، أى تكون قيمة المتوسط الحسابي أكبر من قيمة المنوال ويكون المعامل حينئذ موجباً، والعكس مع التوزيعات الملتوية ناحية اليسار يكون معامل الالتواء سلباً (راجع شكل رقم: ٤-٥، أ، ب، ج).

كما أنه يمكن استخدام المقياس التالي وهو المسمى بمعامل بيرسون Pearson للالتواء:

$$(٢) \text{ معامل بيرسون للالتواء} = \frac{٣ (\text{المتوسط الحسابي} - \text{الوسيط})}{\text{الانحراف المعياري}}$$

$$\frac{٣ (\bar{S} - \bar{O})}{E} = \text{م.ت.١}$$

وما ذكر عن المعامل م.ت.١ يقال أيضاً عن المعامل م.ت.٣ إلا أن قيمة م.ت.٣ تنحصر فيما بين + ١ ، - ١ ، كما أن هذه الصيغة أكثر دقة من صيغة م.ت.١ حيث أن المنوال أقل دقة من الوسيط في وصف البيانات لأنه لا يأخذ في إعتباره إلا القيم الأكثر تكراراً ويهمل باقي القيم.

ويلاحظ على المقياسين السابقين في حساب الالتواء أنهما يعتمدان على المتوسط الحسابي والانحراف المعياري. وكما ذكرنا آنفاً أنه قد يتعذر في بعض الأحيان حساب المتوسط وبصفه خاصة في التوزيعات المفتوحة، كما أن حساب الانحراف المعياري يحتاج بدوره إلى عمليات حسابية طويلة. لكل ذلك استخدمت

فكرة الفرق بين الربيعين (الأعلى والأدنى) والوسيط فى تحديد مقدار الإلتواء (معامل الإلتواء) الذى يمكن حسابه اذن على النحو التالى:

$$\text{معامل الإلتواء} = \frac{(\text{الربيع الأعلى} - \text{الوسيط}) - (\text{الوسيط} - \text{الربيع الأدنى})}{(\text{الربيع الأعلى} - \text{الربيع الأدنى})}$$

أى أن:

$$\text{م.ت.م} = \frac{(\text{ر} - \text{ط}) - (\text{ط} - \text{أ})}{(\text{ر} - \text{أ})}$$

ونظراً لأن الفرق بين قيمة الربيع الأعلى والوسيط يساوى الفرق بين الوسيط والربيع الأدنى فى التوزيعات المتماثلة. فإذا كان هناك فرق بين كل من المقدارين كان ذلك دليلاً على عدم تماثل التوزيع، ووجود بعض القيم الشاذة هى التى تسبب هذا الفرق، وبالتالي التواء المنحنى ناحية اليمين أو اليسار أى التواء المنحنى التواء موجباً أو سالباً. ومن المعلوم أيضاً أنه فى حالة التوزيعات المتماثلة يقع الربيعان على بعدين متساويين من الوسيط، بينما فى التوزيعات الملتوية يختلف بعدا الربيع الأعلى والربيع الأدنى عن الوسيط، وبذلك يكون الفرق بين بعديهما عنه - أيضاً - مقياساً للإلتواء.

ويسمى معامل الإلتواء المحسوب بالصيغة السابقة بمقياس بولى Bowley - ويتميز بأنه المقياس الوحيد الذى يمكن استنتاجه من الرسم بدون اللجوء إلى حساب أى قيمة وذلك باستخدام المنحنى التكرارى المتجمع الصاعد أو النازل. ويعاب عل بالمقياس السابق أنه لا يأخذ فى اعتباره قيم المفردات قبل قيمة الربيع الأدنى وكذلك قيم المفردات بعد قيمة الربيع الأعلى.

ونظراً لأن قيمة الإلتواء فى التوزيعات المتماثلة، كما سبق أن ذكرنا، تساوى صفراً فإن هذه القيمة تتخذ أساساً لتقدير نوع ودرجة حدة أو شدة الإلتواء. فكلما قربت قيمة أى معامل من معاملات الإلتواء الثلاثة السابقة من الصفر، كلما دل ذلك على وجود التواء ولكنه التواء بسيط، أما إذا بعدت قيمة معاملات الإلتواء عن الصفر وقربت من +1 أو -1 فإن ذلك يدل على كبر درجة حدة الإلتواء. فإذا

كانت إشارة معامل الإلتواء إشارة موجبة فإن ذلك يكون دليلاً على وجود التواء موجب يبتعد فيه الطرف الأيمن للمنحنى الممثل للتوزيع عن الطرف الأيسر مما يدل بالتالى على وجود بعض القيم المتطرفة (أو قيم كبيرة) والتي تقع فى إطار الفئات الأخيرة وتركز باقى القيم فى إطار الفئات الأولى للتوزيع. أما إذا كانت إشارة معامل الإلتواء إشارة سالبة فإن ذلك يكون دليلاً على وجود التواء سالب يبتعد فيه الطرف الأيسر لمنحنى التوزيع عن طرفه الأيمن مما نستنتج منه وقوع بعض القيم الشاذة (قيم صغيرة) فى إطار الفئات الأولى وتركز باقى القيم فى إطار الفئات الأخيرة ن التوزيع التكرارى للظاهرة قيد البحث.

مثال: من توزيع تكرارى لمساحة عدد ١٠٠ من المزارع حسبت المقاييس الإحصائية الآتية:

فدان	
٢٧,٣٠	المتوسط الحسابى للتوزيع
٢٧,٢٤	الوسيط
٢٧,٤٢	المتوال
٢٤,٠٠	الربيع الأدنى
٣٤,٧٠	الربيع الأعلى
٩,٨٠	الانحراف المعيارى

وكان المطلوب استنتاج معامل الإلتواء بالطرق السابقة، فإن ذلك ويكون على النحو التالى:

$$\begin{aligned}
 (١) \text{ م.ت} &= \left[\frac{٢٧,٤٢ - ٢٧,٣}{٩,٨} \right] = ٠,٠٠٦ \\
 (٢) \text{ م.ت} &= \left[\frac{(٢٧,٢٥ - ٢٧,٣)^٣}{٩,٨} \right] = ٠,٠١٥ \\
 (٢) \text{ م.ت} &= \left[\frac{(٢٥,٠ - ٢٧,٢٥) - (٢٧,٢٥ - ٣٤,٧)}{(٢٥,٠ - ٣٤,٧)} \right] = ٠,٠١٤
 \end{aligned}$$

وكما هو ملاحظ فإننا حصلنا على نتائج مختلفة لمعامل الالتواء من حيث نوع الالتواء، إلا أنها تتفق جميعاً على وجود التواء ولكنه ضئيل جداً ويدل ذلك على أن التوزيع قريب جداً من التماثل. ويرجع ذلك إلى أن الأساس الذي حسب عليه معامل الالتواء يختلف من طريقه لأخرى، كما أن كل طريقة تلائم بيانات خاصة ولا تلائم سواها. ويجب أن نوجه النظر هنا إلى أنه عند مقارنة التواء توزيعات مختلفة - لابد من استخدام صيغة واحدة لإيجاد معامل الالتواء حتى يكون أساس المقارنة موحداً.

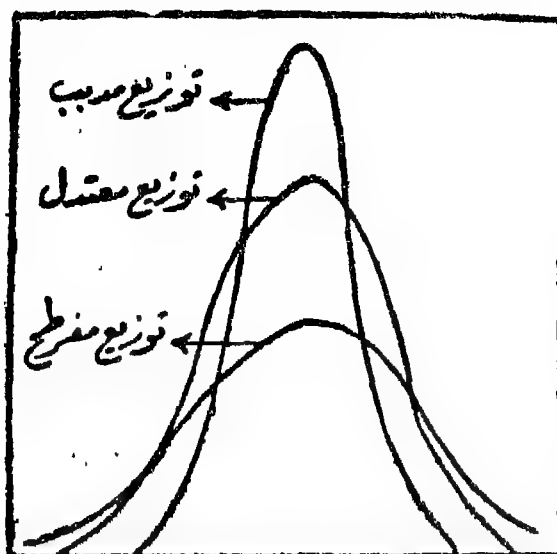
ويعتبر الالتواء مقياساً إحصائياً له أهمية خاصة في مجال الدراسات الجغرافية الكمية لأن معظم المتغيرات الجغرافية التي يمكن قياسها وجمع البيانات عنها تنصف توزيعاتها بأنها توزيعات شديدة الالتواء مما يقف عائقاً أمام تطبيق الاختبارات الكمية البارامترية Parameteric والتي تتطلب أن يكون التوزيع التكراري لبيانات المتغير قيد البحث توزيعاً متماثلاً. هذا من ناحية ومن ناحية أخرى فإننا إذا استخدمنا أحد مقاييس لوصف الأخرى كالمتوسط الحسابي - بصفة خاصة - لوصف توزيع المتغير موضع الدراسة فإنه يكون مضللاً إذا كان بمفرده دون أن يقترن بتوضيح درجة ونوع التواء التوزيع. ولسوق مثلاً لتوضيح ذلك: إذا كان متوسط ما تنفقه الوحدة المحلية الواحدة بمحافظة ما على تنظيم النسل في سنة ١٩٧١ هو ٧,٤٧ جنيهاً لكل ١٠٠٠ من السكان، فإذا اتخذت قيمة هذا المتوسط لذاتها فإنه يمكن افتراض أن نصف عدد الوحدات المحلية - تقريباً - بهذه المحافظة تنفق أكثر من هذه القيمة والنصف الآخر من الوحدات ينفق أقل من ٧,٤٧ جنيهاً على تنظيم النسل بين سكانها. ولكن إذا تبين أن ربع عدد الوحدات هي التي تنفق على هذا الغرض أكثر من المتوسط ٧,٤٧ جنيهاً فإننا نتوقع أن العدد الباقي من الوحدات، وهو أكثر من النصف، ينفق أقل من المتوسط أولاً ينفق شيئاً على تنظيم النسل بين السكان في عام ١٩٧١. وفي مثل هذه الحالة - فإن المتوسط لا يفيدنا كثيراً كمقياس إحصائي يعتمد عليه في استخلاص المعلومات والنتائج. ولكن إذا قمنا

برسم المدرج التكرارى لمثل هذا التوزيع فان ذلك سوف يلقي ضوءاً سريعاً على حقيقة أن هذا التوزيع غير متماثل أو أنه ملتوٍ لتواء شديداً.

التفرطح Kurtosis

لا يقف تحليل المنحنيات البيانية التي تصف الكثير من المظاهر الخاصة ببيانات التوزيعات التكرارية على تحديد أو حساب كل من مقاييس النزعة المركزية أو التشتت أو حتى الالتواء، بل يمتد إلى تحديد تفرطح أو درجة تدبب المنحنيات البيانية الوحيدة القمة. ويعرف التفرطح إحصائياً بأنه ذلك المقياس الذى يقيس الامتداد الذى تتركز فيه القيم Values فى أحد أجزاء التوزيع التكرارى. فمثلاً إذا كانت إحدى فئات التوزيع أو مجموعة من الفئات المتجاورة تحتوى على نسبة كبيرة من تكرارات القيم داخل التوزيع، فإن هذا يعنى أن التوزيع مفرطح بدرجة كبيرة. ولكن درجة تفرطح أو شكل قمة المنحنيات البيانية تختلف من توزيع لآخر. فقد نجد قمة المنحنى البيانى لأحد التوزيعات عريضة أى مفرطحة، وهذا يعكس تركيز القيم فى هذا التوزيع حول متوسطها الحسابى فى مدى كبير. ويسمى التوزيع فى هذه الحالة بتوزيع مفرطح Flat or Platykurtic. وقد نجد أن قمة التوزيع تبدو على شكل أكثر تدبباً، وهذا يعكس صغر مدى تركيز القيم حول المتوسط الحسابى فيظهر شكل المنحنى البيانى ضيق فى الجزء العلوى ومتسع فى الجزء الأوسط. ويسمى التوزيع فى هذه الحالة توزيع مدبب Peaky or Leptokurtic. وقد نجد قمة منحنى التوزيع لا تبدو على شكل مفرطح أو مدبب وهذا يعكس تركيز القيم حول متوسطها الحسابى بدرجة متماثلة. ويسمى التوزيع فى هذه الحالة توزيع متوسط التفرطح (توزيع متماثل) Mesokurtic (شكل رقم ٦-٢).

ويمكن التعرف على تفرطح أو تدبب المنحنيات البيانية بسهولة من خلال الشكل العام لها، غير أن هناك مقياس إحصائى يحدد درجة التفرطح فى التوزيعات بطريقة دقيقة اعتماداً على حساب مجموع القوة الرابعة لانحراف القيم عن متوسطها الحسابى مقسوماً على حاصل ضرب عدد القيم فى القوة الرابعة للانحراف المعيارى. وإذا وضعنا ذلك فى صيغة جبرية فإنها تكون على النحو التالى:



شكل رقم (٦-٢)
أنواع التفرطح لمنحنيات التوزيعات التكرارية

مقياس التفرطح = $\frac{\text{مجم} (س - س')^4}{ن ع^4}$ (٦-٤)

حيث

مجم = مجموع

س = القيم المفردة كل على حدة

س' = المتوسط الحسابي

ن = عدد القيم المكونة لسلسلة البيانات

ع = الانحراف المعياري

وجدير بالذكر أن مقياس التفرطح غير شائع الاستخدام، أو أن إستخدامه ليس كما يجب أن يكون عليه، كمقياس إحصائي لوصف البيانات في مجال الدراسات الجغرافية، على الرغم من الفوائد الهامة التي يمكن الحصول عليها نيم حسابه لمجموعة من البيانات، وهو بذلك يتشابه مع مقياس الإلتواء ومن المفيد أيضاً أن تشير إلى أن كثيراً من المتغيرات الجغرافية تتصف بشدة التواء وتدبب منحنياتها البيانية مما يجعل استخدام مقياس الوصف الإحصائي الأخرى كالمتوسط الحسابي والإنحراف المعياري أقل أهمية لما تعطيه من إنطباعات مضللة عن خصائص توزيع بيانات تلك المتغيرات، ولو كانت هذه البيانات تخص عينات فإنها لا يمكن أن تكون ممثلة أو مسحوبة من مجتمعات إحصائية متماثلة التوزيع، كما إنها تكون غير ملائمة لتطبيق أساليب التحليل البارامترية التي تشترط أن تكون توزيعات البيانات متماثلة.

العزوم ومقياس الإلتواء والتفرطح:

تستخدم العزوم Moments في الإحصاء لبيان تماثل توزيع البيانات وكلمة «عزم» مشتقة من علم الاستاتسكا الذي يوضح أن قدرة القوة على تحريك جسم ما حول محور تتوقف على عاملين هامين هما: مقدار القوة، وبعد القوة عن المحور. وعلى ذلك فإن عزم القوة أو العزم حول المحور يعرف رياضياً على أنه حاصل ضرب مقدار القوة في طول ذراعها (الذراع هو بعد خط عمل القوة عن مركز العزم أو البعد العمودي بين القوة وبين المحور). وتكون عزوم مجموعة من القوى مساوية مجموع حاصل ضرب كل قوة في ذراع عزمها. ومقياس العزم في الإحصاء يختلف على قياسه في الاستاتيسكا (علم الثوابت) حيث يمكن اعتبار التكرارات في التوزيعات ماثلة للقوة والقيمة المناظرة لهذه التكرارات (الفئة في التوزيع) مطابقة للذراع العزم.

والمفهوم الإحصائي للعزم يتطلب تحديد النقطة التي نحسب عندها العزم. فقد يحسب العزم مثلاً حول الصفر أو حول المتوسط الحسابي أو حول أى وسط فرضي آخر. إلا أن حساب العزوم حول المتوسط الحسابي (س) أصبح متعارفاً عليه كنقطة تحسب عندها العزوم.

وسنعرض فيمايلي الصيغ الجبرية لحساب العزم حول المتوسط الحسابي للبيانات المبوبة.

$$١- \text{العزم الأول (م) حول المتوسط الحسابي} = \frac{1}{\text{مجم ك}} (\text{م} - \text{س}) \text{ ك}$$

ولكن مجموع انحرافات القيم عن متوسطها الحسابي $(\text{م} - \text{س}) = \text{صفرًا}$

$$\therefore \text{العزم الأول (م) حول المتوسط الحسابي} = \frac{1}{\text{مجم ك}} (\text{م} - \text{س}) \text{ يساوى صفرًا}$$

$$٢- \text{العزم الثاني (م)} \text{ حول المتوسط الحسابي} = \frac{1}{\text{مجم ك}} (\text{م} - \text{س}) \text{ ك}^٢, \text{ع}^٢ = \text{التباين.}$$

$$٣- \text{العزم الثالث (م)} \text{ حول المتوسط الحسابي} = \frac{1}{\text{مجم ك}} (\text{م} - \text{س}) \text{ ك}^٣, \text{وحيث أن م} = \text{مركز الفتحة, (م} - \text{س}) = \text{الانحراف عن المتوسط الحسابي}$$

$$\text{ح فإن} \\ \text{العزم الثالث (م)} = \left(\frac{\text{مجم ح ك}^٣}{\text{مجم ك}} \right) - ٣ \left(\frac{\text{مجم ح ك}^٢}{\text{مجم ك}} \times \frac{\text{مجم ح ك}}{\text{مجم ك}} \right) + ٢ \left(\frac{\text{مجم ح ك}}{\text{مجم ك}} \right) (٦ - ٨)$$

$$٤- \text{العزم الرابع (م)} \text{ حول المتوسط الحسابي} = \frac{1}{\text{مجم ك}} (\text{م} - \text{س}) \text{ ك}^٤ \\ (٦ - ٩) \dots\dots\dots$$

أو

$$\text{(م)} = \left(\frac{\text{مجم ح ك}^٤}{\text{مجم ك}} \right) - ٤ \left(\frac{\text{مجم ح ك}^٣}{\text{مجم ك}} \times \frac{\text{مجم ح ك}}{\text{مجم ك}} \right) + ٦ \left(\frac{\text{مجم ح ك}^٢}{\text{مجم ك}} \right) \times \left(\frac{\text{مجم ح ك}}{\text{مجم ك}} \right) - ٣ \left(\frac{\text{مجم ح ك}}{\text{مجم ك}} \right) (٦ - ١٠) \dots\dots\dots$$

ويمكن الاستطرداف بنفس الطريقة السابقة للحصول على العزم الأعلى حول المتوسط الحسابى، ولكننا عادة لانحتاج فى الدراسات الجغرافية العملية، وبصفة خاصة عند تحليل الرواسب المفتتة، إلى عزم أعلى من العزم الرابع.

وتجدر الإشارة هنا إلى أن العزم الثالث فى صورته السابقة (معادلة ٦-٨) يمكن اعتباره مقياساً دقيقاً للإلتواء وحيث أن قيمة هذا العزم تساوى صفر للتوزيعات المتماثلة. أو بمعنى آخر إذا كان العزم الثالث يساوى صفراً فإن الانحرافات السالبة تكون مساوية للانحرافات الموجبة. وفى هذه الحالة فإن قيمة الإلتواء تساوى صفراً. أما فى التوزيعات غير المتماثلة فقد تكون قيمة العزم الثالث سالبة وهذا يعنى وجود التواء فى التوزيع ناحية اليسار، أما إذا كانت قيمته موجبة فإن هذا يعنى وجود إلتواء فى التوزيع ناحية اليمين. وكلما كانت قيمة العزم الثالث قريبة من الصفر كلما كان منحنى التوزيع قريباً من التماثل، أما إذا كانت قيمته كبيرة (موجبة أو سالبة) كان المنحنى ملتوياً بشدة.

ولما كان العزم الثالث مقياساً بمعكبات الوحدات المعيارية (القياسية) الأصلى فلا بد فى حساب الإلتواء عن نسبة هذا العزم إلى أحد مقاييس التشتت أو الاختلاف مثل التباين (مربع الانحراف المعيارى) للتوزيع بعد تكعيب الأخير وعلى ذلك فإن:

$$\frac{(\text{العزم الثالث})^2}{(\text{العزم الثانى})^3} = \frac{(\text{العزم الثالث})^2}{(\text{التباين})^3} = \text{معامل الإلتواء}$$

$$\text{م.ت.} = \frac{(\mu_3)^2}{(\mu_2)^3} \dots \dots \dots (16 - 1)$$

كما تستخدم فكرة لعزم الرابع لقياس درجة التفرطح فى التوزيعات بطريقة دقيقة وذلك بقسمة العزم الرابع حول المتوسط الحسابى للتوزيع على مربع الانحراف المعيارى له، ويمكن تحويل ذلك إلى صيغة إحصائية على النحو التالى:

$$\frac{\text{العزم الرابع}}{\text{العزم الثاني}^2} = \frac{\text{العزم الرابع}}{\text{التباين}^2} = \text{درجة التفرطح}$$

$$\frac{42}{(22)^2} = \dots\dots\dots (12 - 6)$$

ويعزى السبب فى استخدام العزم الرابع (الذى يحتاج إلى رفع انحرافات القيم عن متوسطها الحسابى إلى القوة الرابعة) إلى وجود بعض القيم المتطرفة التى تمثل التوزيع، أما إذا لم توجد هذه القيم فإن العزم الرابع يعطى نتائج منخفضة، وبالقيمة على مربع العزم الثانى فإن قيمة التفرطح ستكون منخفضة أيضاً كما يرجع السبب فى القسمة على مربع العزم الثانى إلى التخلص من وحدات القياس والحصول على مقياس نسبى، وبحساب قيمة التفرطح للتوزيعات المتماثلة وجد أنها تساوى (٣) واعتبر التوزيع المتماثل متوزع متوسط التفرطح أما إذا كان مقدار التفرطح أقل من (٣) فإن منحنى التوزيع يعد متفرطحاً منخفض القمة، بينما إذا كانت قيمة التفرطح أكبر من (٣) فإن ذلك يعكس وجود منحنى مدبباً يرتفع عن مستوى المنحنى المتماثل (شكل رقم ٦-٢). وفى النوع الأخير من المنحنيات تتركز القيم بشدة حول المتوسط الحسابى للتوزيع.

وتطبيقاً لما سبق يمكن حساب كل من قيمتى الإلتواء والتفرطح عن طريق استخدام العزوم من الجدول التالى (جدول رقم ٦-١) الذى يبين التوزيع التكرارى لمائة من الحيازات الزراعية فى أحد المراكز الإدارية حسب المساحة لهذه الحيازات

مساحات	التكرارات	مراكز	ح	(د/ح)	(د/ح) ك	(د/ح) ك ^٢	(د/ح) ك ^٣
- ٥	٣	٧,٥	٢٠-	- ٤	١٢-	٤٨	- ١٩٢-
- ١٠	٩	١٢,٥	١٥-	- ٣	٢٧-	٨١	- ٢٤٣-
- ١٥	١٣	١٧,٥	١٠-	- ٢	٢٦-	٥٢	- ١٠٤-
- ٢٠	١٦	٢٢,٥	٥-	- ١	١٦-	١٦	- ١٦-
- ٢٥	٢٠	٢٧,٥	صفر	صفر	صفر	صفر	صفر
- ٣٠	١٥	٣٢,٥	٥	١	١٥	١٥	١٥
- ٣٥	١٣	٣٧,٥	١٠	٢	٢٦	٥٢	٣٠١
- ٤٠	٨	٤٢,٥	١٥	٣	٣٤	٧٢	٢١٦
- ٤٥	٣	٤٧,٥	٢٠	٤	١٢	٧٤	١٩٢
الاجمعي	١٠٠				٨١-	٣٧٨	٥٥٥+
					٧٧+		٨٢٥+
					- ٤		- ٧٨

ح = الإنحراف عن المتوسط (م - س)

ل = طول الفئة

ويكون حساب كل من العزوم الثلاثة حول المتوسط الحسابي كما يلي:

١- حساب العزم الثاني حول المتوسط الحسابي:

$$م_2 = \frac{1}{ك} \text{مجم} \left(\frac{ح}{ل} \right)^2 ك$$

$$= \frac{1}{100} \times 384 = 3.84 \text{ من الوحدات المربعة}$$

ولما كانت الوحدة المستخدمة (طول الفئة) هي ٥ أفدنة

$$\text{فإن } م_2 = 3.84 \times (5)^2 = 96$$

٢- حساب العزم الثالث حول المتوسط الحسابي :

$$م_3 = \frac{\text{مجم} \left(\frac{ح}{ل} \right)^3 ك}{\text{مجم} ك} \times 3 - \frac{\text{مجم} \left(\frac{ح}{ل} \right)^2 ك}{\text{مجم} ك} \times 3 + \frac{\text{مجم} \left(\frac{ح}{ل} \right) ك}{\text{مجم} ك} \times 2 =$$

$$= \frac{28}{100} \times 3 - \frac{143}{100} \times 3 + \frac{4}{100} \times 2 = \frac{40}{100} \times 2 = 0.8$$

$$= 0.1807 \text{ من الوحدات المكعبة}$$

ولما كانت الوحدة المستخدمة (طول الفئة) هي ٥ أفدنة

$$\text{فإن } م_3 = 0.1807 \times (5)^3 = 22.59$$

٣- ويكون العزم الرابع حول المتوسط الحسابي هو :

$$\begin{aligned}
 & \frac{\text{مجد (ح/ل) ك}}{\text{مجد ك}} \times \frac{\text{مجد (ح/ل) ك}^2}{\text{مجد ك}} \times \frac{\text{مجد (ح/ل) ك}^4}{\text{مجد ك}} = 4^2 \\
 & \left(\frac{\text{مجد (ح/ل) ك}}{\text{مجد ك}} \right)^3 - \left(\frac{\text{مجد (ح/ل) ك}}{\text{مجد ك}} \right)^2 \times \frac{\text{مجد (ح/ل) ك}^2}{\text{مجد ك}} \times 6 + \\
 & \left(\frac{4 -}{100} \right)^3 - 2 \left(\frac{4 -}{100} \right) \left(\frac{384}{100} \right) \times 6 + \frac{4 -}{100} \times \frac{28 -}{100} \times 4 - \frac{3360}{100} =
 \end{aligned}$$

= 33,59 من وحدات ذات القوة الرابعة (الأس الرابع)

$$4(5) \times 33,59 =$$

$$20993,75 =$$

وبناء على النتائج السابقة يمكن حساب قيمة كل من إلتواء ودرجة التفرطح للتوزيع كمايلي:

$$\frac{2(22)}{3(22)} = \frac{2(\text{العزم الثالث})}{3(\text{العزم الثالث})} = \text{معامل الإلتواء} = (1)$$

$$\frac{2(22,59)}{3(96)} = \text{م.ت} = 4$$

$$0,0006 =$$

وهذا يعنى أن إلتواء التوزيع موجب وضعيف جداً، أى أنه توزيع قريب جداً من التماثل أو الاعتدال كما هو راضح من البيانات فى الجدول.

$$2 - \text{درجة التفرطح} = \frac{\text{العزم الرابع}}{2(\text{العزم الثانى})} = \frac{42}{22} =$$

$$\frac{20993,75}{2(96)} =$$

$$2,279 =$$

وهذا يعنى أن التوزيع مفرطح ولكنه بدرجة لا تختلف كثيراً من تفرطح التوزيع المتماثل إذا أن قيمة التفرطح التى حصلنا عليها قريبة من ٣ (درجة تفرطح التوزيع المتماثل).

الباب الثالث

التقدير الإحصائي وأساليب المقارنة

Statistical Estimation and Comparison Techniques

مقدمة

الفصل السابع: تقدير خصائص (معالم) المجتمع

الفصل الثامن: اختبارات الفروض الاحصائية

الفصل التاسع: أساليب المقارنة الباراميتريّة (المعلمية)

الفصل العاشر: أساليب المقارنة غير الباراميتريّة (غير المعلمية)

الباب الثالث

التقدير الإحصائي وأساليب المقارنة

مقدمة:

بعد دراستنا لأهم المقاييس الكمية المستخدمة في وصف البيانات الجغرافية باستخدام أسلوب المعاينة، ودراسة بعض التوزيعات الاحتمالية لبعض المجتمعات الهامة لإلقاء الضوء على المفاهيم الأساسية والأساليب النظرية لتحديد احتمال صفات مفردات البيانات، وحيث أنه يتعذر في كثير من الأحيان دراسة جميع مفردات المجتمع ولكن يمكن من الناحية العملية دراسة مفردات عينة مسحوة من هذا المجتمع، فاهتمامنا الآن ينصب أولاً على مشكلة هامة في الاستدلال الإحصائي Statistical Inference وهي مشكلة تقدير معالم المجتمع (مثل المتوسط الحسابي التباين ...) المجهولة من إحصائيات العينة المعروفة.

وقد تتطلب المعالجة الإحصائية في دراسة ظاهرة ما أن يضع الباحث بعض الفروض عن الخصائص (المعالم) المجهولة لمجتمع هذه الظاهرة، ويختار من بينها الفرض الذي لا يكون معلوماً لديه، كما يحدد البديل المقابل للفرض قيد الاختبار. فمثلاً إذا أراد الباحث تقدير المتوسط العام لمجتمع ظاهرة ما عليه أولاً أن يفترض قيمة معينة لهذه المعلمة المجهولة ثم يسحب عينة من بين مفردات المجتمع ليختبر ما إذا كانت الإحصائية المحسوبة من بيانات العينة تتفق أو لا تتفق مع فرضه السابق عن

معلمة المجتمع. واختبار مدى صحة الفرض أو قياس درجة الثقة في التقدير تستخدم الاحتمالات ويسمى هذا الأسلوب الاحصائي باختبار الفروض Test of Hypotheses أو ما يعرف وأحياناً باختبار الدلالة أو المعنوية Test of Significance. وبما أن جوهر البحث الجغرافي هو بيان، أو التعرف على الفروق والاختلافات المكانية Spatial differentiation فإنه يمكن الاستفادة والاستعانة بأساليب المقارنة الاحصائية الكمية للإجابة على كثير من التساؤلات التي تتصل بمدى دلالة أو أهمية هذه الفروق والاختلافات.

وفي هذا الباب سنتناقش الأساليب الكمية الباراميتريّة (المعلمية) والأساليب اللاباراميتريّة (اللامعلمية) من خلال استخدامها للمقارنة بين خصائص مفردات مجموعة واحدة من البيانات وتوزيع نظري لها، وللمقارنة بين مفردات مجموعتين أو بين ثلاث مجموعات أو أكثر من البيانات. وتجدد الإشارة هنا إلى أن صلاحية أو تطبيق أسلوب معين من هذه الأساليب الكمية يتوقف على نوع البيانات (نوعية، ترتيبية إلخ) المقيسة للظاهرة قيد البحث، وعلى خصائص الأسلوب الكمي نفسه.

الفصل السابع

تقدير خصائص (معالم) المجتمع

تقدير خصائص (معالم) المجتمع

رأينا في الفصول السابقة أن دراسة المجتمعات تعتمد أساساً على الحصر الشامل لجميع مفردات المجتمع للتعرف على خصائص (معالم) هذا المجتمع. ويقصد بالمجتمع في الدراسات الكمية، كما سبق أن ذكرنا، كل المفردات التي يتكون منها هذا المجتمع والتي تتصف بوحدة أو أكثر من الصفات المميزة المشتركة، وأن أية قيمة تحسب من توزيع المجتمع لدراسة خصائصه تسمى معلمة Parameter فالتوسط الحسابي، التباين، والانحراف المعياري هي معالم لهذا المجتمع Population parameters.

وكما ذكرنا أن دراسة المجتمعات عن طريق أخذ كل مفردات المجتمع تعتبر من الأمور غير اليسيرة التي تحتاج إلى وقت طويل ومجهود كبير. هذا إلى جانب أنه في كثير من الأحيان تتعذر دراسة كل المجتمع إذ يكون حجم المجتمع غير محدود Infinite أي لا يمكن حصر جميع مفرداته، مثل مجتمع إنتاج سلعة من نوع معين، مما يفرض على الباحث القيام بفحص جزء من هذا المجتمع أو «عينة» منه. كذلك قد تتعذر دراسة كل المجتمع المحدود finite أي الذي يمكن حصر جميع مفرداته وذلك لأسباب اقتصادية أو عملية تقف أمام اتباع أسلوب الحصر الشامل لمعرفة القيمة الحقيقية لمعلمة المجتمع. فمثلاً إذا أردنا معرفة متوسط عمر المصباح من إنتاج أحد المصانع في فترة معينة (المجتمع في هذه الحالة هو مجتمع محدود ويتكون من الكمية المنتجة من المصابيح) فإنه يتعين علينا إضاءة كل مصباح من

إنتاج المصنع حتى يحترق لمعرفة عمره وبذلك نتمكن من معرفة متوسط عمر المصباح في المصنع كله. أى أنه لمعرفة معلمة مجتمع المصابيح نضطر إلى اتلاف جميع مفرداته وهذا غير ممكن عملياً كما أنه يكون مكلف اقتصادياً. لذلك فإنه من الأوفق أن تأخذ عينة من هذه المصابيح الكهربائية وتترك مضاءة حتى تحترق ثم نستخدم قيمة متوسط عمر المصابيح في العينة (احصائية العينة) كتقدير غير متحيز لمتوسط المجتمع (معلمة المجتمع). نخلص من ذلك أنه فى كثير من الأحيان لانستطيع معرفة القيمة الحقيقية بصفة مؤكدة لمعلمة المجتمع قيد البحث عن طريق الحصر الشامل ولذا فإننا نلجأ إلى تقديرها عن طريق اختيار عينة عشوائية من هذا المجتمع وحساب قيمة تقديرية لهذه المعلمة من بيانات العينة.

أنواع التقدير:

هناك نوعان من التقديرات لمعالم المجتمع هما: تقدير النقطة وتقدير الفترة الذى قد يكون مصحوباً بدرجة الثقة فى صحة التقدير. ويعبر تقدير النقطة Point Estimate عن معلمة المجتمع بقيمة واحدة كقيمة المتوسط الحسابى لعينة الذى يتخذ كتقدير غير متحيز أو تقدير قريب جداً من المتوسط الحقيقى للمجتمع. أما تقدير الفترة Interval Estimate أو فترة الثقة Confidence Interval فيعبر عن مدى معين من القيم بحيث يشتمل هذا المدى على قيمة المتوسط العام (أو المعلمة) للمجتمع. ولتوضيح ذلك نذكر أنه إذا قيس مسافة على خريطة فكانت ٥٢,٨ سenti متر أو إذا قدر عمر قطعة أثرية بألف سنة فإننا فى هذه الحالة بصدد تقدير نقطة لطول المسافة ولعمر القطعة الأثرية. ومن الناحية الأخرى إذا ذكرنا أن المسافة هى $52,8 \pm 3$ سenti متر أى أن المسافة تقع بين ٥٢,٥ سenti متر و ٥٣,١ سenti متر باحتمال قدره ٠,٩٥، أو إذا قدر عمر قطعة أثرية بين ١٠٠٠ سنة و ١٥٠٠ سنة باحتمال قدره ٠,٩٥، فإننا نعطي تقديراً بفترة لطول المسافة أو لعمر القطعة الأثرية واحتمال أن يقع كل منهما فى أى نقطة منها. ونظراً لأن تقدير الفترة يعطى قيمة احتمال وقوع (٠,٩٥) وكذلك عدم وقوع (٠,٠٥)، طول المسافة

أو العمر في هذه الفترة، ولذلك فإننا نطلق على الفترة (٥, ٥٢, ١, ٥٣) سنتيمتراً أو ١٠٠٠, ١٥٠٠ سنة) إسم «فترة ثقة ٩٥%» لطول المسافة ولعمر القطعة الأثرية. وبصفة عامة فإن التقديرات بفترة تشير إلى معنوية أو دقة التقدير، وبالتالي فإنها تفضل على التقدير بنقطة لمعلمة من معالم المجتمع.

تقدير فترة الثقة لمعالم المجتمع:

كثيراً ما يكون هناك مجتمع لاتعرف معالمه (المتوسط الحسابي «م» أو الانحراف المعياري «ع») ونجد أنه بينما نريد معرفة بعض أو كل هذه المعالم فإننا لانستطيع تحديد هذه المعالم تحديداً دقيقاً ومؤكداً وذلك لأسباب عملية أو اقتصادية. وفي هذه الحالة نلجأ إلى تقدير معالم المجتمع الأصلي من خلال إحصائيات عينة عشوائية تسحب من هذا المجتمع. على أنه يمكن القول أن هذا التقدير يتطلب معرفة طبيعة العلاقة بين التوزيع الأصلي للمجتمع بمعالمه المختلفة وتوزيع المعاينة للمتوسطات الحسابية بإحصائياته المختلفة.

ومن الناحية العملية لا يمكننا سحب عدد كبير من العينات، لاعتبارات مالية وزمنية، ولكن يمكن سحب عينة واحدة ونحسب منها المتوسط الحسابي (س) ومنه يمكن تقدير المتوسط الحسابي للمجتمع (م) بإنشاء فترة ثقة للمتوسط الحسابي المحسوب من العينة (س). فإذا كان حجم العينة كبيراً بدرجة كافية (ن أكبر من ٣٠) ومسحوبة عشوائياً من مجتمع - ليس بالضرورة أن يكون معتدلاً - له متوسط حسابي (م) وانحراف معياري (ع) فإننا يمكن أن نتوقع، تبعاً لخصائص التوزيع المعتدل، أن نجد قيمة فعلية للمعلمة (م) تقع بالتقريب:

في الفترة $[س - ع, س + ع]$ باحتمال ٦٨, ٢٧%

وفي الفترة $[س - ٢ع, س + ٢ع]$ باحتمال ٩٥, ٤٥%

وفي الفترة $[س - ٣ع, س + ٣ع]$ باحتمال ٩٩, ٧٣%

أما إذا رجعنا إلى خصائص توزيع المعاينة للمتوسطات الحسابية المسحوبة

عشوائياً من المجتمع المراد تقدير المعلمة (م) له، فإننا يمكن أن نتوقع تبعاً لخصائص التوزيع العيني، الذى يتوزع بشكل تقريبي كالتوزيع المعتدل الذى متوسط حسابى م س وانحراف معيارى (خطأ معيارى) $\frac{\sigma}{\sqrt{n}}$ ، أن نجد أن قيمة المعلمة (م) تقع بالتقريب:

$$\text{فى الفترة } \left[\bar{S} - \frac{\sigma}{\sqrt{n}}, \bar{S} + \frac{\sigma}{\sqrt{n}} \right] \text{ باحتمال } 68,27\%$$

$$\text{وفى الفترة } \left[\bar{S} - \frac{\sigma^2}{\sqrt{n}}, \bar{S} + \frac{\sigma^2}{\sqrt{n}} \right] \text{ باحتمال } 95,45\%$$

$$\text{وفى الفترة } \left[\bar{S} - \frac{\sigma^3}{\sqrt{n}}, \bar{S} + \frac{\sigma^3}{\sqrt{n}} \right] \text{ باحتمال } 99,73\%$$

وتسمى الفترات 68,27%، 95,45%، 99,73% بفترات الثقة، أو مستوى الثقة Confidence level، (الاحتمال) لتقدير المعلمة (م)، وبالتالى فإن درجة عدم الصحة فى التقدير للفترة الثانية = (100 - 95,45)% = 4,55% وعدم للصحة فى التقدير للفترة الثالثة = (100 - 99,73)% = 0,27% كما يطلق على حدود هذه الفترات (س ± ع، س ± 2 ع، س ± 3 ع) اسم حدود الثقة Confidence limits للمعلمة (م). أما حدود الثقة للفترات 95، 99 فهى على الترتيب: س ± 1,96 ع، س ± 2,58 ع. وتسمى القيم 1,96، 2,58 فى حدود الثقة بمعاملات الثقة Confidence Coefficients أو القيم الحرجة Critical Values ويرمز لها بالرمز (ز). ويمكن الحصول على معاملات الثقة من مستوى الثقة المطلوب للتقدير، أو العكس، كما يلى:

مستوى الثقة (الاحتمال)	68,27%	80%	90%	95%	96%	98%	99%
(ز)	1,00	1,28	1,645	1,96	2,00	2,33	2,58

من كل مما سبق يمكن القول أن إنشاء فترة الثقة يعتمد أساساً على الخصائص التي سبق ذكرها عن التوزيع الاحتمالي المعتدل والتوزيع الاحتمالي للمتوسطات الحسابية للعينات الذي سبق أن قلنا أن له توزيع معتدل إذا كان حجم العينة كبير بدرجة كافية بصرف النظر عن شكل التوزيع الأصلي للمجتمع المسحوب منه العينة، وأن المتوسط الحسابي للمتوسطات الحسابية لجميع العينات الممكنة يساوي تقريباً المتوسط العام (م)، كما أن الانحراف المعياري للمتوسطات الحسابية يساوي الانحراف المعياري للمجتمع مقسوماً على حجم العينة (مع إهمال معامل التصحيح أو «معامل بسل» Bessel's Correction في العينات الكبيرة) أو ما يعرف بالخطأ المعياري Standard Error وهو الخطأ الناجم عن احتمال بعد أو قرب خصائص العينة في تمثيل معالم المجتمع، أي أنه عبارة عن مدى تفاوت متوسط العينة مثلاً عن متوسط المجتمع المسحوبة منه. وبما أن الانحراف المعياري للمجتمع ثابت بينما حجم العينة متغير فإن ذلك يعني أنه كلما زاد حجم العينة كلما صغرت قيمة الخطأ المعياري، وكلما كان الخطأ المعياري صغيراً كلما كانت قيمة متوسط العينة قريبة من قيمة متوسط المجتمع، وكان بالتالي تمثيل على معرفة الانحراف المعياري للمجتمع فيمكن تقديره بسهولة. ولكن قد يحدث في بعض الأحيان أن لا يكون الانحراف المعياري للمجتمع معلوماً. وللتغلب على هذه الصعوبة فإننا نلجأ إلى تقدير الخطأ المعياري عن طريق استعاضة الانحراف المعياري للمجتمع بالانحراف المعياري للعينة.

وقد جرت العادة في الأبحاث الجغرافية على إنشاء فترة بدرجة ثقة ٩٥٪ إلا أنه يمكن أيضاً إنشاء فترة بأية درجة ثقة. ونظراً لعدم التأكد من أن فترة الثقة المحسوبة تشتمل أو لا تشتمل على المتوسط الحقيقي (المعلمة) للمجتمع فإن تقديرنا ينشأ على أساس عنصر الاحتمال إذ يمكن أن نتحكم في نسبة الخطأ الذي قد تقع فيه باستخدام درجة ثقة معينة في التقدير، فإذا استخدمنا درجة ثقة كبيرة فإن فترة الثقة المرتبطة بها تكون كبيرة. وبالطبع كلما زاد طول فترة الثقة كلما

قلت قيمتها العملية لذلك فإنه من المهم أن تكون فترة الثقة التي نقررها ذات فائدة عملية للبحث.

التقدير من احصائية (مقاييس) العينات:

ذكرنا من قبل أنه إذا كان لدينا مجتمعاً، ليس بالضرورة أن يكون توزيعه معتدلاً، متوسطة (م) وانحرافه المعياري (ع) وسحبنا منه كل العينات الممكنة التي حجمها (ن) فإن توزيع المعاينة للمتوسطات الحسابية للعينات يقترب من التوزيع المعتدل الذي متوسطه (م) وانحرافه المعياري (الخطأ المعياري) $\frac{\sigma}{\sqrt{n}}$. وكما نعلم

من نظرية النهاية المركزية التي مؤداها أنه إذا كان لدينا عينة حجمها كبير بدرجة كافية (ن أكبر من ٣٠) ومسحوبة عشوائياً من مجتمع (ليس بالضرورة أن يكون معتدلاً تماماً) له متوسط حسابي (م) وانحراف معياري $\frac{\sigma}{\sqrt{n}}$ فإن قيمة المتوسط

الحسابي للعينة (س) يمكن اعتبارها تقديراً غير متحيز لمتوسط (معلمة) المجتمع إذا ربطنا هذا التقدير بدرجة معينة للثقة أو بنسبة معينة للخطأ في التقدير والتي على أساسها ننشئ فترة الثقة للمتوسط الحسابي المحسوب من العينة بدرجة ثقة ٩٥% أو ٩٩% (بمعنى أن احتمال أن تشتمل هذه الفترة المقدرة على المتوسط العام للمجتمع (م) يساوي ٩٥، أو ٩٩)، لا بد أن تمتد هذه الفترة بين ١,٩٦+، - ١,٩٦ أو ٢,٥٨+، - ٢,٥٨ من الانحراف المعياري للمتوسطات الحسابية أو من الخطأ المعياري، ولذلك فإن:

حدود فترة الثقة للمتوسط الحسابي للمجتمع (م) =

$$s \pm z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \quad \dots \quad \dots \quad \dots \quad (1 - \alpha)$$

في حالة ما إذا كانت العينة مسحوبة من مجتمع غير محدود أو إذا كانت

المعاينة بارجاع من مجتمع محدود حيث (س + ز $\alpha'_{1/2}$ $\times \sqrt{\frac{\sigma^2}{n}}$) هو الحد الأعلى لفترة الثقة، (س - ز $\alpha'_{1/2}$ $\times \sqrt{\frac{\sigma^2}{n}}$) هو الحد الأدنى لفترة الثقة. أما إذا كانت المعاينة بدون ارجاع من مجتمع محدود حجمه ن فإن:
حدود فترة الثقة للمعلمة (م) =

$$س \pm ز_{\alpha'_{1/2}} \times \sqrt{\frac{\sigma^2}{n}} \quad \dots \quad \dots \quad \dots \quad \sqrt{\frac{n-n-1}{1-n}} \quad \dots \quad (2-10)$$

وفي الحالات التي لا يكون فيها الانحراف المعياري (ع) معلوماً يستعاض عنه بالانحراف المعياري للعينة (ع) للحصول على حدود الثقة السابقة لتقدير متوسط المجتمع ويكون ذلك صحيحاً إذا كان حجم العينة أكبر من ٣٠ مفردة.
- مثال (١)

إذا كان توزيع الأجور لعمال أحد المصانع يتوزع توزيعاً قريباً جداً من الاعتدال، وبأخذ عينة عشوائية حجمها ١٠٠ عامل من عمال هذا المصنع وجد أن متوسط الأجور في العينة هو ٧٠ جنيهاً في الشهر فأوجد فترة ثقة ٩٥٪ للمتوسط الحسابي لأجور العمال في هذا المصنع علماً بأن الانحراف المعياري لأجور العمال في المصنع هو ١٠ جنيهاً.
بما أن الانحراف المعياري للمجتمع معلوماً فإن:

$$١ = \frac{١٠}{\sqrt{١٠٠}} = \frac{\sigma}{\sqrt{n}} = \text{الخطأ المعياري}$$

ومن جداول التوزيع المعتدل المعياري نجد أن المساحة تحت المنحنى التي تشمل على ٩٥٪ (درجة الثقة) من القيم تنحصر بين ز $\alpha_{1/2}$ (أى ١٠ - ٩٥،
= ٠,٥ ÷ ٢ = ٠,٢٥) $\pm ١,٩٦$.

$$\therefore \text{الحد الأدنى لفترة الثقة } 95\% = \bar{S} - 1,96 \times \sqrt{\frac{E}{n}}$$

$$= 70 - 1,96 \times 1 = 68,04 \text{ جنيهاً}$$

$$\text{الحد الأدنى لفترة الثقة } 95\% = \bar{S} + 1,96 \times \sqrt{\frac{E}{n}}$$

$$= 70 + 1,96 \times 1 = 71,96 \text{ جنيهاً}$$

وبذلك تكون (68,04)، (71,96) هي فترة الثقة 95% لمتوسط المجتمع (المعلمة م).

مثال ٢:

أرادت مصلحة الضرائب بمحافظة الإسكندرية معرفة متوسط الأرباح التجارية السنوية للمحلات الصغيرة لتقدير الضرائب المستحقة على أصحاب هذه المحلات فسحبت عينة عشوائية من 100 محل حسب منها المتوسط الحسابي للمبيعات السنوية فكان 1100 جنيهاً كما حسب الانحراف المعياري للأرباح مجتمع المحلات التجارية فكانت 120 جنيهاً والمطلوب إيجاد فترة ثقة 99% لمتوسط الأرباح لمجتمع هذه المحلات.

بما أن الانحراف المعياري للمجتمع معلوماً فإن:

$$\therefore \text{الخطأ المعياري} = \sqrt{\frac{E}{n}} = \sqrt{\frac{120}{100}} = 12 \text{ جنيهاً}$$

وحيث أن أكبر من 30 فإننا نستخدم جداول المنحنى المعتدل المعياري لإيجاد القيمة المقابلة لقيمة $z_{\alpha/2}$ فتكون:

$$z_{\alpha/2} = 2,58 = (1 - 0,99) \div 2 = 0,01 \div 2$$

$$\therefore \text{الحد الأدنى لفترة } 99\% = \bar{s} - 2,58 \times \frac{e}{\sqrt{n}}$$

$$= 1100 - 12 \times 2,58 = 1069,04 \text{ جنيهها}$$

$$\text{الحد الأعلى لفترة الثقة } 99\% = \bar{s} + 2,58 \times \frac{e}{\sqrt{n}}$$

$$= 1100 + 12 \times 2,58 = 1130,96 \text{ جنيهها}$$

إذا نتوقع أن متوسط الأرباح للمحلات التجارية الصغيرة في محافظة الاسكندرية يقع في الفترة بين ١٠٦٩٠,٠٤، ١١٣٠,٩٦ جنيه بدرجة ثقة ٩٩٪. مثال (٣) :

لمعرفة متوسط استدارة الرواسب الحصوية على جزء من الشاطئ سحبت عينة عشوائية من هذه الرواسب حجمها ١٠٠ حصوة وحسب منها المتوسط الحسابي (٥٠ ملليمتر) والانحراف المعياري (١٠ ملليمتر). والمطلوب حساب فترة الثقة ٩٥٪ لتقدير المتوسط العام لمجتمع الرواسب الحصوية على الشاطئ قيد الدراسة.

بما أن الانحراف المعياري لمجتمع الرواسب الحصوية على الشاطئ غير معلوم، وأن حجم العينة أكبر من ٣٠ مفردة فإننا نحسب الخطأ المعياري للتقدير باستخدام الانحراف المعياري للعينة :

$$\text{الخطأ المعياري} = \frac{e}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1$$

وباستخدام جداول المنحنى المعتدل المعياري نجد أن:

$$1,96 \pm = (a)_{1/2} z = 2 \div 0,05 = 0,95 - 1 = \alpha_{1/2}$$

∴ الحد الأدنى لفترة الثقة ٩٥% = $1 \times 1,96 - 50 = 48,04$ ملليمتر

الحد الأعلى لفترة الثقة ٩٥% = $1 \times 1,96 + 50 = 51,96$ ملليمتر

وهذا يعنى أن متوسط استدارة الرواسب الحصوية على الشاطئ يقع بين الحدين (٤٨,٠٤، ٥١,٩٦ ملليمتر)، أو بمعنى آخر لا يقل عن ٤٨,٠٤ ملليمتر ولا يزيد عن ٥١,٩٦ ملليمتر بدرجة ثقة ٩٥% ونسبة خطأ مسموح به ٥%.

وفى بعض الأحيان يرغمنا عنصر التكاليف فى جمع البيانات إلى سحب عينة صغيرة (ن أقل من ٣٠ مفردة) لتقدير معالم المجتمع الذى تمثله. وفى مثل هذه الحالات لا يمكن الاستعانة بنظرية النهاية المركزية إذ أن توزيع للقيم المعيارية للمجتمع غير معلوماً والعينة حجمها صغير فإنه يمكن استبداله بالانحراف المعيارى للعينة بعد إجراء بعض التعديل على الأخير حتى نحصل على «أحسن تقدير للانحراف المعيارى للمجتمع» لأن الخطأ المعيارى المحسوب من واقع الانحراف المعيارى للبيانات المشاهدة من العينة الصغيرة قد يختلف كثيراً عن الانحراف المعيارى للمجتمع الأصلي مما يؤثر على درجة الدقة فى التقدير والاستنتاج الاحصائى، وبذلك فإن:

أحسن تقدير للانحراف المعيارى للمجتمع (ع) = الانحراف المعيارى للعينة
(د) × تصحيح «بل»، أى أن:

$$\hat{c} = c \times \sqrt{\frac{n}{n-1}}$$

ويكون الخطأ المعيارى للمتوسط الحسابى (س) عبارة عن :

$$\frac{\hat{c}}{\sqrt{n}} = \text{الخطأ المعيارى}$$

$$\hat{\sigma} = \sqrt{\frac{\hat{\sigma}^2}{n-1}}$$

وتسمى القيمة (ن - ١) بعدد المتغيرات المستقلة الخطية التي يمكن تكوينها من ن من القيم المساهدة أو ما يعرف بدرجات الحرية.

وفى حالة العينات الصغيرة أيضاً تستخدم الإحصائية ت (١) بدلاً من (ز) لتقدير القيمة الافتراضية لمعلمة المجتمع (م)، وذلك لأن الأولى تتميز بأن توزيعها ينتشر على مدى أوسع من التوزيع المعتدل، ومن ذلك نتوقع أن تحتاج إلى أكثر من خطأين معيارين لتحديد فترة الثقة ٩٥٪ لمتوسط المجتمع (١,٩٦) فى حالة التوزيع المعتدل المعيارى (ز)، وتعتمد قيمة (ت) على حجم العينة أو بالأحرى على درجات الحرية (ن - ١) ولذلك فإن المتغير المعيارى (ت) ليس له توزيع إ احتمال واحد كالمتغير المعيارى (ز)، ولكن له توزيع إ احتمالى لكل قيمة من قيم درجات الحرية (من ١ إلى ٣٠). والجدول المختصر فى ملاحق الكتاب يبين قيمة (ت) عند مستويات معنوية مختلفة لمساحة الطرف الموجب للمنحنى وهو كاف لإستخدامه فى إنشاء فترات الثقة للعينات الصغيرة. فإذا كان حجم العينة = ١٠، ومستوى المعنوية (α) = ٠,٠٥ فإن قيمة (ت) التى تجعل مساحة كل طرف من طرفى المنحنى = ٢,٥٪ من المساحة الكلية تقع فى الصف تحت درجة الحرية ٩ (أى ١٠-١) وفى العمود ٠,٠٢٥ تساوى ٢,٢٦٢ أى أن ٩٥٪ من المساحة الكلية للمنحنى تنحصر بين (ت) + ٢,٢٦٢ (لاحظ أن القيمة المناظرة فى التوزيع المعتدل المعيارى (ز) = ± ١,٩٦). إلا أنه عندما يكبر حجم العينة تصبح قيمة $\hat{\sigma}$ قريبة جداً من σ ، كما يقترب التوزيع الاحتمالى للمتغير (ت) من التوزيع الاحتمالى للمتغير الاحتمالى للمتغير المعتدل المعيارى (ز)، وهكذا يمكن استخدام التوزيع المعتدل كتقريب لتوزيع (ت) إذا كانت درجات الحرية تساوى أو أكثر من ٣٠. (ويلاحظ من جدول توزيع «ت» أن القيمة الأخيرة عند «ن - ١» = α هى نفسها قيمة «ز» فى المنحنى المعتدل). وبذلك يقتصر استخدام توزيع (ت) على

الحالات التي تكون فيها حجم العينة أقل من ٣٠ والانحراف المعياري للمجتمع غير معلوم.

ويمكن إنشاء فترة ثقة لتقدير متوسط المجتمع (م) بنفس الطريقة المستخدمة سابقاً في حالة العينات الكبيرة، إلا أننا نستخدم في هذه الحالة قيم (ت) بدرجات حرية (ن - ١) بدلاً من قيمة (ز). وتكتب صيغة حدود فترة الثقة لمعلمة المجتمع (م) كالتالي:

$$\text{حدود فترة الثقة} = \bar{m} \pm t \times \frac{\hat{\sigma}}{\sqrt{n-1}} \quad \dots \quad \dots \quad \dots \quad (3-10)$$

مثال (٤):

أخذت عينة من مجموعة من الروافد ذات الرتبة الأولى في أحد الأحواض النهرية مكونة من ٢٥ رافداً لدراسة انحدار جوانبها فوجد أن متوسط الانحدارات هو ٢٠ درجة بانحراف معياري ٥ درجات، والمطلوب تقدير متوسط انحدار جوانب كل الروافد من نفس الرتبة وذلك بدرجة الثقة ٩٥٪.

نحسب أولاً أحسن تقدير للانحراف المعياري للمجتمع (ع) وهو يساوي:

$$\hat{\sigma} = c \times \frac{\bar{c}}{\sqrt{n-1}} = ٥ \times \frac{٢٥}{\sqrt{٢٤}} = ٥,١$$

$$\text{ويكون الخطأ المعياري للعينة} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{٥,١}{\sqrt{٢٥}} = ١,٠٢$$

وحيث أن الانحراف المعياري للعينة هو المعروف فتستخدم في هذه الحالة قيمة (ت) المناظرة لدرجة الثقة ٩٥٪. ولدرجة الحرية (٢٥ - ١). ومن جدول (ت) يظهر أن هذه القيمة تساوي ٢,٠٦٤.

∴ متوسط انحدارات جمع روافد الرتبة الأولى = متوسط الانحدارات فى العينة \pm قيمة (ت) \times الخطأ المعيارى.

$$\therefore \text{متوسط انحدارات جميع الروافد} = 20 \pm 2,064 \times 1,02$$

$$= 2,11 \pm 20$$

أى أن هناك احتمال مقداره ٩٥% أن يكون متوسط انحدارات جميع الروافد النهرية من الرتبة الأولى فى هذا الحوض النهري عبارة عن قيمة تتراوح بين ١٧,٨٩ درجة و ٢٢,١١ درجة.

مثال (٥):

فى المثال رقم (٢) إذا رأيت مصلحة الضرائب أن التكلفة المخصصة لفحص الإيرادات السنوية للمحلات الصغيرة فى محافظة الإسكندرية لا تكفى لدراسة عينة كبيرة، فسحبت عينة عشوائية حجمها ٢٦ محلاً، ولنفرض أن المتوسط الحسابى لإيرادات هذه المحلات هو أيضاً ١١٠٠ جنيه والانحراف المعيارى المحسوب من بيانات هذه العينة هو ١٢٠ جنيه أيضاً. والمطلوب إنشاء فترة ثقة بدرجة الثقة ٩٥%.

بما أن حجم العينة صغيراً (ن أقل من ٣٠) والانحراف المعيارى للعينة معلوماً، فإننا نستخدم قيمة (ت) بدرجات الحرية (ن - ١) = ٢٥ التى تجعل طرفى المنحنى تساوى ١ - ٩٥ = ٠,٠٥، هى ٢,٠٦. والخطأ المعيارى لهذه العينة يحسب له قيمة أحسن تقدير للانحراف المعيارى للمجتمع وهى:

$$\hat{\sigma} = \frac{n}{1-n} \sqrt{\frac{n}{1-n}} \times \sigma$$

$$= \frac{26}{1-26} \sqrt{\frac{26}{1-26}} \times 120 =$$

$$122,4 = 1,02 \times 120 =$$

$$24,48 = \frac{122,4}{25} \sqrt{\frac{\hat{\sigma}}{n}} = \text{الخطأ المعياري}$$

ويكون الحد الأدنى لفترة الثقة ٩٥٪ = ١١٠٠ - ٢,٠٦ × ٢٤,٤٨

$$= 105,17 \text{ جنيهاً.}$$

الحد الأعلى لفترة الثقة ٩٥٪ = ١١٠٠ + ٢,٠٦ × ٢٤,٢٨

$$= 1149,3 \text{ جنيهاً.}$$

أى إذا كررنا هذه النسبة بعدد كبير جداً من المرات فإننا نتوقع أن معلمة المجتمع (م) تقع قيمتها بين ١١٤٩,٨٣، ١٧، ١٠٥٠، جنيهاً فى ٩٥٪ من الحالات.

وبمقارنة حدى فترة الثقة السابقة بمثيلتها التى سبق تقديرها لهذا المثال فى حالة حجم العينة الكبير وباستخدام التوزيع المعتدل المعيارى نجد أن فترة الثقة للعينة الصغيرة أكبر من فترة الثقة للعينة الكبيرة لأن منحني التوزيع (ت) أكثر تفرطحاً من منحني التوزيع المعتدل وذلك لأنه يأخذ فى اعتباره الخطأ الناشئ من تقدير الانحراف المعياري المحسوب من بيانات العينة الصغيرة. وبصفة عامة إذا كان الانحراف المعياري للمجتمع (ع) معلوماً فإن فترة الثقة الخاصة بعدد كبير من العينات من نفس الحجم لها مدى معين ثابت بالرغم من اختلاف قيمة مراكزها

(المتوسطات الحسابية للعينات) وذلك لأن الخطأ المعياري $\left[\left(\frac{\sigma}{n} \right) = \text{خ.م} \right]$

مقدارها ثابت لكل العينات. أما إذا كانت قيمة الانحراف المعياري للمجتمع (ع) غير معلومة فإن الخطأ المعياري (خ.م) فى هذه الحالة لا يكون مقدار ثابتاً بل سيختلف من عينة لأخرى وتبعاً لذلك فإن فترة الثقة للمتوسطات الحسابية المختلفة

المحسوبة من عينات ذات حجم متساوى لها مراكز مختلفة ومدى مختلف أيضاً. ويوضح ذلك الجدول التالى (جدول رقم: ١٠ - ١) والذي يعتمد على بيانات خاصة بالمثال رقم (٣).

جدول رقم (١٠ - ١)

العلاقة بين الانحراف المعياري والخطأ المعياري وفترة الثقة

أ- فى حالة معرفة قيمة (ع)					
ن	م	ع	مخ	حدود الثقة	فترة (مدى) الثقة
				٢٩٥	٢٩٥
١٠٠	٥٠	٢٠	٢,٠٠	٥٤,٠٠ ، ٤٦,٠٠	٨,٠٠
١٠٠	٦٠	٢٠	٢,٠٠	٦٤,٠٠ ، ٥٦,٠٠	٨,٠٠
١٠٠	٧٠	٢٠	٢,٠٠	٧٦,٠٠ ، ٦٤,٠٠	٨,٠٠
١٠٠	٨٠	٢٠	٢,٠٠	٨٨,٠٠ ، ٧٢,٠٠	٨,٠٠
ب- فى حالة عدم معرفة قيمة (ع)					
١٠٠	٥٠	١٠	١,٠٠	٥٢,٠٠ ، ٤٨,٠٠	٤,٠٠
١٠٠	٦٠	٢٠	٢,٠٠	٦٤,٠٠ ، ٥٦,٠٠	٨,٠٠
١٠٠	٧٠	٣٠	٣,٠٠	٧٦,٠٠ ، ٦٤,٠٠	١٢,٠٠
١٠٠	٨٠	٤٠	٤,٠٠	٨٨,٠٠ ، ٧٢,٠٠	١٦,٠٠

التقدير من نسبة العينة:

كثيراً ما تواجه الباحث الجغرافى بعض الحالات التى لا يمكن فيها قياس المفردات المشاهدة ولكن يمكن حصر المفردات المشاهدة التى لها خاصية معينة، فمثلاً يمكن حصر العمال الإناث من الذكور فى صناعة النسيج أو حصر

المساحات المزروعة أذرة من المزروعات الصيفية على المستوى القومى. وعموماً فإن مفردات المجتمع يمكن أن تنقسم إلى أكثر من قسمين حسب طبقاً للصفات أو الخصائص المراد دراستها، كأن تنقسم مفردات المجتمع السكانى - حسب الحرف - إلى مفردات حرفتها الزراعة وثانية حرفتها الصناعة وأخرى حرفتها التجارة ... إلخ. ولكن فى بعض الأحيان، قد يتكون المجتمع من مجموعتين أو قسمين متميزين أحدهما له صفة أو خاصية معينة والآخر ليس فيه هذه الصفة أو الخاصية. وبعبارة أخرى يمكن أن نقسم مفردات المجتمع إلى وحدات موجبة وأخرى سالبة طبقاً للخاصية المراد إختبارها ودراستها، وتكون الوحدات الإيجابية هى الوحدات التى تتصف بهذه الخاصية بينما لا تتصف الوحدات السلبية بهذه الخاصية. فمثلاً يمكن تقسيم مجتمع الذكور فى سن معينة إلى أميين ومتعلمين، أو تقسيم مجتمع المواليد إلى أطفال ذكور وإناث، أو تقسيم مجتمع إنتاج إحدى الآلات إلى إنتاج معيب وآخر غير معيب ... إلخ. وقد يهمنا أحياناً أن نعرف نسبة كل مجموعة أو قسم (أى المفردات التى تمتلك الصفة المراد دراستها) فى المجتمع. ولكنه فى معظم الأحيان لا يمكن قياس أو تحديد نسبة الصفة التى تتصف بها مفردات مجتمع ما فى المجتمع كله عن طريق الحصر الشامل لسبب من الأسباب التى شرحناها سابقاً، وعليه فإننا نقوم بسحب عينة عشوائية من هذا المجتمع ونحدد منها نسبة المفردات التى تمتلك الصفة التى نريد دراستها، وتأخذ قيمة النسبة فى العينة كمقدر نقطة غير متحيز لنسبة المجتمع. فمثلاً إذا كانت نسبة الأمية المحددة من بيانات عينة مسحوبة عشوائياً من مجتمع محافظة ما هى $Q = 0.05$ ، فى هذه الحالة يمكن اعتبار أن نسبة الأمية فى هذه المحافظة كلها (د) هى 0.05 وهنا يمكن أن نعتبر أن Q هى تقدير نقطة غير متحيز للمعلمة (د)

وبصفة عامة إذا افترضنا أن نسبة عدد مفردات المجتمع التى تتصف بهذه الصفة هى (ق) وكانت النسبة (ق) قريبة جداً من الصفر أو الواحد الصحيح فإن التوزيع الاحتمالى للنسبة المحسوبة من البيانات المشاهدة فى عينة حجمها (ن) كبير

نسبياً ويقترب من التوزيع المعتدل الذى متوسطه الحسابى يساوى (د) وتباينه هو $\frac{d(d-1)}{n}$. ومعنى ذلك أن المتوسط الحسابى لقيم (ق) المحسوبة من كل العينات المختلفة (توزيع المعاينة للنسب) الممكنة المتساوية الحجم يساوى (د)، كما أن تباين توزيع (ق) يقل إذا كبر حجم العينة (ن). وبذلك إذا كانت (ن) كبيرة بدرجة كافية فإن قيم (ق) تتركز حول (د)، أى أن تشتتها حول المتوسط نادراً ما يكون بمقدار كبير.

ويمكن تقدير فترة النسبة فى مجتمع باستخدام عينة كبيرة الحجم (ن أكبر من ٣٠) عن طريق المعادلة الآتية:

$$\text{تقدير فترة النسبة} = ق \pm ز \frac{1}{2} \alpha \times \text{مق}$$

حيث ق هى النسبة فى العينة، خ. مق هى الخطأ المعيارى للنسبة. وتكون بذلك حدود الثقة للنسبة فى المجتمع كما يلى:

$$\text{حدود الثقة للنسبة} = ق \pm ز \frac{1}{2} \alpha \times \sqrt{\frac{ق(١-ق)}{ن}} \dots (١٠ - ٤)$$

وذلك فى حالة إذا كانت المعاينة من مجتمع محدود أو إذا كانت المعاينة بإرجاع من مجتمع محدود أما إذا كانت المعاينة بدون إرجاع من مجتمع محدود (حجمه ن) فإن:

$$\text{حدود الثقة بالنسبة فى المجتمع} =$$

$$ق \pm ز \frac{1}{2} \alpha \times \sqrt{\frac{ق(١-ق)}{ن}} \times \sqrt{\frac{ن-١}{١-ق}} \dots (١٠ - ٥)$$

وتسمى القيمة $\sqrt{\frac{ق(١-ق)}{ن}}$ بالخطأ المعيارى لتوزيع إحصائية نسبة العينة

(ق).

مثال (٦):

سحبت عينة عشوائية من ٢٠٠ أسرة (حجم كل منها ٥ أفراد) من سكان منطقة معينة لمعرفة رأى هذه الأسر فى تطبيق أسلوب جديد لتنظيم النسل، فوجد أن ١٢٠ أسرة تستخدم الأسلوب المراد تطبيقه. قدر بدرجة الثقة ٩٥٪ نسبة الأسر المستخدمة للأسلوب الجديد لتنظيم النسل فى هذه المنطقة.

نسبة الأسر المستخدمة للأسلوب الجديد لتنظيم النسل (ق)

$$٠,٦٠ = \frac{١٢٠}{٢٠٠} = \frac{\text{عدد الأسر}}{\text{حجم العينة}} =$$

$$١ - ق = ١ - ٠,٦ = ٠,٤$$

$$٩٥\% = \alpha - ١٠٠ = \alpha$$

$$٢,٥\% = \alpha^{١/٢}, \quad ٥\% = \alpha$$

$$\therefore ١,٩٦ \pm \alpha^{١/٢}$$

الحد الأعلى لفترة الثقة ٩٥٪ للأسر المستخدمة للأسلوب الجديد لتنظيم النسل

$$\frac{٠,٤ \times ٠,٦}{٢٠٠} \sqrt{\quad} \times ١,٩٦ + ٠,٦ =$$

$$٠,٦٦٨ = ٠,٦٨ + ٠,٠٨$$

الحد الأدنى لفترة الثقة ٩٥٪ للأسر المستخدمة للأسلوب الجديد لتنظيم

النسل

$$\frac{٠,٤ \times ٠,٦}{٢٠٠} \sqrt{\quad} \times ١,٩٦ - ٠,٦ =$$

$$٠,٥٣٢ = ٠,٦٨ - ٠,١٤٨$$

وعلى ذلك فإن نسبة الأسر المستخدمة للأسلوب الجديد لتنظيم النسل فى هذه

المنطقة يقع بين (٠,٥٣٢ ، ٠,٦٦٨) وذلك بدرجة ثقة ٩٥٪ ونسبة خطأ ٥٪.

مثال (٧):

فى استطلاع للرأى العام بالعينة سحبت عينة عشوائية حجمها ١٠٠ من جميع الناخبين فى حى معين بإحدى المدن حيث دلت على أن أصوات ٥٥٪ منهم ستكون فى صالح مرشح معين، أوجد حدود الثقة ٩٥٪، ٩٩٪، ٩٩,٧٣٪ للنسبة بين جميع الناخبين المؤيدين لهذا المرشح. وما هو حجم العينة التى يجب أخذها من الناخبين بحيث يكون ٩٥٪، ٩٩,٧٣٪ منهم واثقين من أن هذا المرشح سوف يختار من مرشحين اثنين.

حدود الثقة ٩٥٪ لنسبة مجتمع الناخبين

$$\begin{aligned} & \frac{\sqrt{q(1-q)}}{n} \sqrt{1,96 \pm q} = \\ & \frac{\sqrt{0,45 \times 0,55}}{100} \sqrt{1,96 \pm 0,55} = \\ & 0,10 \pm 0,55 = \end{aligned}$$

حدود الثقة ٩٩٪ لنسبة مجتمع الناخبين

$$\begin{aligned} & \frac{\sqrt{0,45 \times 0,55}}{100} \sqrt{2,58 + 0,55} = \\ & 0,13 \pm 0,55 = \end{aligned}$$

حدود الثقة ٩٩,٧٣٪ لنسبة مجتمع الناخبين

$$\begin{aligned} & \frac{\sqrt{0,45 \times 0,55}}{100} \sqrt{3 \pm 0,55} = \\ & 0,15 \pm 0,55 = \end{aligned}$$

ويكون حجم العينة المطلوبة بدرجة الثقة ٩٥٪ هو:

$$\text{حجم العينة} = q \pm z \alpha^{1/2} \times \frac{\sqrt{q(1-q)}}{n}$$

$$\frac{0,45 \times 0,55}{100} \sqrt{z_{\alpha}^{1/2} \pm 0,55} =$$

$$\frac{\alpha_{z,50}}{n} \sqrt{\pm 0,55} =$$

وحيث أننا استخدمنا التقدير $0,55 = q$ على أساس البيانات السابقة، وبما أن المرشح سينجح فقط إذا حصل على أكثر من 50% من أصوات مجتمع الناخبين، فإنه يجب أن تكون القيمة $\frac{\alpha_{z,50}}{n} \sqrt{\pm 0,55}$ أقل من $0,5$

$$\frac{\alpha_{z,50}}{n} \sqrt{\pm 0,55} = 0,5 \text{ هو: } 95\% \text{ حجم العينة لدرجة الثقة}$$

$$\frac{1,96 \times 0,5}{n} \sqrt{\pm 0,55} = 0,5$$

$$n = 384,2 \text{ (} 385 \text{ ناخباً على الأقل)}$$

$$\frac{\alpha_{z,99,73}}{n} \sqrt{\pm 0,55} = 0,5 \text{ هو: } 99,73\% \text{ حجم العينة لدرجة الثقة}$$

$$\frac{3 \times 0,5}{n} \sqrt{\pm 0,55} = 0,5$$

$$n = 900 \text{ ناخباً على الأقل.}$$

الفصل الثامن
اختبارات الفروض الإحصائية
Testing of Hypotheses

الفصل الثامن .

اختبارات الفروض الإحصائية

رأينا في الفصل السابق كيف يمكن الاعتماد على توزيعات المعاينة لإيجاد فترة الثقة لبعض معالم المجتمع المجهولة. وفي هذا الفصل سندرس بعض اختبارات الفروض الإحصائية الهامة المبنية على أساس المتوسطات الحسابية والتباين للعينات وسنجد أن هناك صلة وثيقة بين التقدير الإحصائي والاختبار الإحصائي. وفي اختبارات الفروض الإحصائية تواجهنا مشكلة اتخاذ قرار بقبول فرض معين أو رفضه، ويتم اتخاذ هذا القرار بناء على البيانات التي نحصل عليها من عينة. فمثلاً إذا قلنا أن متوسط كمية الأمطار في الاقليم (أ) يساوي متوسط كمية الأمطار في الاقليم (ب) فإننا نطرح بذلك فرض يحتمل الصواب والخطأ: بمعنى أن هناك احتمال أن يكون متوسط كمية الأمطار متشابهاً في الاقليمين. ويتم اتخاذ قرار بقبول أو رفض هذا الفرض بعد أخذ عينة من كميات الأمطار في فترة محددة وحساب متوسطيهما للاقليمين، وذلك لأنه من الصعب كما عرفنا جمع البيانات عن مجتمع كميات الأمطار بأسلوب الحصر الشامل، أي بشكل دقيق، لذا فإنه يجب أن يكون اختيار العينة صحيحاً حتى تكون النتائج النهائية متشابهة إلى حد كبير النتائج التي يمكن الحصول عليها لو استخدمنا بيانات المجتمع كله. ولما كانت النتائج التي تستقى من اختبار أية عينة غير ممثلة تمثيلاً كاملاً أو غير مطابقة

تماماً لنتائج المجتمع فإن الفرض الإحصائي الخاص بمجتمع ما هو قول يحتمل الصواب والخطأ ولا بد من جمع مجموعة من البيانات لمعرفة مدى انطباق صحة أو عدم صحة هذا الفرض على النتائج المتحصل عليها. فإذا كانت النتائج تتفق مع الفرض يقبل الفرض وبالتالي يمكن تعميمه. أما إذا لم تتفق النتائج مع الفروض يرفض الفرض. ويتم قبول أو رفض باستخدام الأساليب الإحصائية الكمية التي تتيح للباحث إتخاذ القرار المناسب في ظل ظروف التشكك في عدم التأكد.

ولاشك أن إختبارات الفروض واتخاذ قرار بشأنها يعد من أصعب الأمور. وللتسهيل في عرض أسلوب التحليل الكمي ستعرض فقط للمشكلات التي تشتمل على فرضين لاتخاذ قرار بتفضيل أحدهما على الآخر ، وذلك بعد تطبيق القواعد الرئيسية لاختبار هذه الفروض.

قواعد اختبار الفروض الإحصائية:

يمكن تحديد الأسس والقواعد اللازمة لإجراء إختبارات الفروض الإحصائية على النحو التالي:

- ١- وضع الفروض (فرض العدم والفرض البديل).
- ٢- تحديد مستوى المعنوية (مستوى الدلالة).
- ٣- تحديد التوزيع النظري (الاحتمالي) للإحصائية المختبرة.
- ٤- استخدام بيانات العينة لحساب قيمة إحصائية الاختبار واستخدام التوزيع النظري لإتخاذ القرار الإحصائي الخاص برفض أو قبول فرض العدم عن طريق تحديد المنطقة الحرجة (منطقة الرفض) ومنطقة القبول تحت منحني التوزيع النظري. وفيمايلي مناقشة تفصيلية لكل قاعدة من القواعد السابقة.

وضع الفروض:

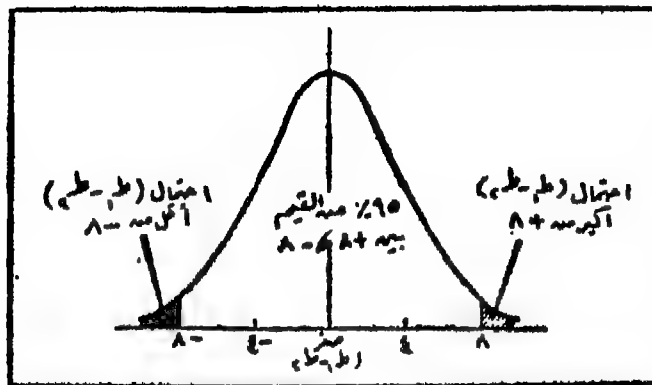
إن أولى الخطوات لإجراء اختبار الفرض هو التعبير عنه رياضياً (أى وضع افتراض معين للمعلمة المراد دراستها ثم يختبر هذا الافتراض فى مقابل المقياس المحسوب من البيانات المشاهدة من العينة) فإذا أردنا اختبار مدى تفوق الأقليم (أ) على الاقليم (ب) فى كمية الأمطار فإن الفرض المناسب فى هذه الحالة هو أن نفترض أن متوسطات كمية الأمطار متساوية فى الاقليمين أى أن متوسط كمية الأمطار فى الاقليم الأول (ط_١) يساوى متوسط كمية الأمطار فى الاقليم الثانى (ط_٢) ، ويصبح الفرض المختبر Testing Hypothesis كمايلى :

$$\text{الفرض المختبر: } \mu_1 = \mu_2$$

ويمكن التعبير عن هذا الفرض بصورة أخرى بأن نقول أن الفرق بين متوسطى كمية الأمطار فى الاقليمين يساوى صفراً، أى أن الفرض ينص على عدم وجود فرق بين المتوسطين (ط_١ - ط_٢ = صفر). ويسمى الفرض فى هذه الحالة بفرض العدم Null Hypothesis ويرمز له بالرمز (H₀) ، وهو الفرض الذى لا يتفق مع البيانات المشاهدة.

فإذا قبل فرض العدم فإن ذلك يعنى أن النتائج جاءت مؤيدة له، أما إذا رفض الفرض فمعنى ذلك أن النتائج لم تكن مؤيدة له، ولذا فإننا نضطر إلى البحث عن الفرض البديل Alternative Hypothesis ويرمز له بالرمز (H₁) وفى المثال بين أيدينا يكون الفرض البديل هو أن متوسطى كمية الأمطار فى الاقليمين غير متساويين، أى أن ط_١ ≠ ط_٢ . ويعرف هذا النوع من الفرض البديل ، الذى ينص على وجود فرق بين المتوسطين بالفرض ثنائى الطرف أو ثنائى الجهة - أى أنه فرض غير محدد Non- Directional . أما إذا توفر للباحث من الأدلة ما يجعله يعتقد بأنه فى حالة عدم تساوى المتوسطين فإن المتوسط الأول يتفوق على المتوسط

الثانى أو أن المتوسط الثانى يتفوق على المتوسط الأول، أى أن $\mu_1 < \mu_2$ أو $\mu_1 > \mu_2$ ، فإن هذا الفرض يعرف بالفرض احادى الطرف أو احادى الجهة - أنه فرض محدد Directional. وتمثل الحالة الأولى فى النصف الأيمن من منحنى التوزيع العينى، بينما تتمثل الحالة الثانية فى النصف الأيسر من المنحنى كما نرى فى الشكل رقم (٨-١).



شكل رقم (٨-١)

لتوزيع العينى للاختلاف بين المتوسطات مقدراً بالقيم المعيارية

فمثلاً إذا كان الفرق أكبر من μ من μ من الوحدات المعيارية فإنه لا يمكن منه تحديد ما إذ كانت μ_1 أكبر من أو أقل من μ_2 ولكن كما نرى أن هذا الفرق يتمثل (يقع) فى المساحة تحت طرفى التوزيع (المساحات المظللة فى الرسم) والتي تشتمل على ٥٪ (٠,٠٥) من احتمال تكرارات هذا الفرق والتي تتوزع على أساس ٢,٥٪ من المساحة الكلية تحت المنحنى (أو احتمال ٠,٢٥) فى الطرف الأيمن من منحنى التوزيع، ومثل هذا المقدار فى الطرف الأيسر. ومن الناحية الأخرى إذا كان الفرض البديل محدداً أى إذا كان الفرق موجباً ($\mu_1 < \mu_2$) أو سالباً ($\mu_1 > \mu_2$) أو سالباً ($\mu_1 > \mu_2$) أو سالباً ($\mu_1 < \mu_2$)، فإنه يمكن تحديد نصف منحنى التوزيع

الذى يوافق هذا الفرق أو الذى تقع فيه قمة الفرض البديل . وكقاعدة عامة إذا كان الفرض البديل محدداً فإن اختبار هذا النوع من الفروض يسمى إختباراً من طرف واحد One-Tailed Test ، أما إذا كان الفرض البديل غير محدد فإن إختباره يسمى الاختبار ثنائى الطرف Two-tailed tests .

مما سبق يمكن أن نستنتج أن هناك أربع حالات لقبول أو رفض فرض العدم (أو رفض أو قبول الفرض البديل) هى كمايلي:

١- أن يكون فرض العدم صحيحاً وأن تؤيد نتائج اختبار العينة صحته، أى يقترب المقياس الاحصائى المحسوى من العينة من المقياس الاحصائى النظرى، وفي هذه الحالة يقبل فرض العدم ويكون قرار القبول صائباً.

٢- أن يكون فرض العدم صحيحاً ولكن لا تؤيد نتائج العينة صحته، فتكون المحصلة هى رفض هذا لفرض، وبذلك يكون هناك خطأ فى الحكم على فرض العدم برفضنا له بينما هو فى الواقع فرض صحيح وقبولنا للفرض البديل وهو فرض غير صحيح. ويعرف الخطأ فى هذه الحالة بالخطأ من النوع الأول Type I error ويرمز له بالرمز α ، أى أنه احتمال رفض فرض العدم بالرغم من أنه فى الواقع صحيح.

٣- أن يكون فرض العدم غير صحيح بينما تأتى العينة بما يثبت ذلك، وتكون المحصلة هو قبول فرض العدم وبذلك يكون هو قبول فرض العدم وبذلك يكون هناك خطأ فى قرار القبول لفرض العدم وهو فى الواقع غير صحيح ورفض الفرض البديل وهو فرض صحيح، ويعرف هذا النوعين من الخطأ بالخطأ من النوع الثانى Type II Error، ويرمز له بالرمز β أى أنه احتمال قبول فرض العدم بالرغم من أنه فى الواقع غير صحيح.

٤- أن يكون فرض العدم غير صحيح ولكن لا تأتى نتائج العينة بما يثبت ذلك، وتكون المحصلة هى رفض فرض العدم وهو فى الواقع غير صحيح، وقبول الفرض البديل وهو فى الواقع صحيح، وبذلك يكون القرار برفض فرض العدم فى هذه الحالة سليماً.

وعليه يتضح لنا أنه عند قبول أو رفض فرض العدم فإننا نتعرض لنوعين من الأخطاء. ويمكن تلخيص القرارات الممكنة السابقة والأخطاء الناجمة عنها في الجدولين التاليين:

جدول رقم (٨-١)
حالات قبول أو رفض فرض العدم

النتيجة	القرارات الممكنة	نوع الفرض	
		الفرض البديل	فرض العدم
القرار صائب	قبول فرض العدم	غير صحيح	صحيح
القرار خاطئ	رفض فرض العدم	غير صحيح	صحيح
القرار خاطئ	قبول فرض العدم	صحيح	غير صحيح
القرار صائب	رفض فرض العدم	صحيح	غير صحيح

جدول رقم (٨-٢)
أنواع أخطاء القرارات رفض أو قبول فرض العدم

الواقع		نوع الفرض
فرض العدم غير صحيح	فرض العدم صحيح	فرض العدم
صائب (β)	خطأ من النوع الأول (α)	رفض
خطأ من النوع الثاني (β)	Type 1 Error	قبول
Type 11 Error	صائب (١-٥)	

اختبار المعنوية (الدلالة) Test of Significance

يعتمد تحديد قيم التوزيعات النظرية لاحصائيات العينات (المعايير الاحصائية) على نسبة أو احتمال الخطأ المسموح به لقبول أو رفض الفروض الاحصائية. وتعرف نسبة الخطأ أو الاحتمال المسموح به بمستوى المعنوية أو مستوى الدلالة

Level of Significance الذى يكون اختباراه فى الواقع الخطوة التالية على طريق إختبارات الفروض الإحصائية. وعند تحديد مستوى المعنوية يجب الأخذ فى الاعتبار نوعى الخطأ فى رفض العدم وهو صحيح، وقد عرفنا ذلك بالخطأ من النوع الأول. فمثلاً إذا قررنا قبول حدوث خطأ من النوع الأول فى خمس مرات كل مائة مرة فإن قرارنا هذا يعنى أنه فى هدد كبير من التجارب نتوقع أن نرفض فرض العدم وهو فى الواقع صحيح ٥٪ من المرات وبذلك يكون الحد الأقصى الذى قررنا قبوله لاحتمال (a) وتسمى قيمة الاحتمال (a) بمستوى المعنوية أو مستوى الدلالة. من هذا نرى أن مستوى المعنوية هو احتمال حدوث خطأ من النوع الأول، أى احتمال رفض فرض العدم وهو فى الواقع صحيح. وبالتالي فإن القيمة التى نحددها لهذا الاحتمال تعتبر الأساس فى الحكم على وجود فروق جوهرية، من الناحية الإحصائية Statistically Significance، بين إحصائيات العينات وبين معالم المجتمع أو ارجاع هذه الفروق إلى الصدفة. وقد جرت العادة على إختبار $\alpha = 0.05$ أو 0.01 فإذا كانت $a = 0.05$ ورفضنا فرض العدم فإننا نستنتج أن نتيجة العينة تختلف جوهرياً عن فرض العدم بمستوى معنوية 0.05 ومن الناحية الأخرى قد تؤدي نتائج العينة إلى قبول فرض العدم وهو فى الواقع غير صحيح، فنكون قد وقعنا فى خطأ من النوع الثانى، ويرمز لاحتمال حدوث هذا الخطأ بالرمز (B). وبالتالي فإن احتمال رفض فرض العدم وهو غير صحيح يساوى $(1-B)$ ويسمى هذا الاحتمال بقوة الاختبار Power of the Test وتسمى قيمة الاحتمال (B). بمستوى الثقة Level of Confidence، وهو يعكس مستوى الدلالة حيث يشير إلى الخطأ فى قبول فرض العدم وهو غير صحيح. وقد جرت العادة على إختبار $\beta = 0.95$ أو 0.99 . فمثلاً إذا كانت $\beta = 0.95$ وقبلنا فرض العدم فإن قرارنا هذا يعنى أنه إذا تكررت التجربة عدد كبير من المرات نتوقع أن نقبل فرض العدم وهو فى الواقع يفر صحيح 0.95 من المرات.

وتجدر الإشارة هنا إلى أن قيمة (α) يتم تحديدها بافتراض صحة فرض العدم بينما تحسب قيمة (β) بافتراض صحة الفرض البديل. كما أنه من الممكن تقليل احتمال الخطأ β ولكن ذلك يكون على حساب زيادة احتمال الخطأ (α)، أما إذا أردنا تقليل احتمال الخطأ (β) بينما يظل احتمال الخطأ (α) ثابت فإنه يجب زيادة حجم العينة، فكلما زاد حجم العينة كلما انخفضت قيمة الانحراف المعياري

وأصبحت خصائص العينة أكثر تمثيلاً لمعالم المجتمع الذى سحبت منه. ونظراً لأنه قد جرت العادة على تحديد مستوى المعنوية (α) قبل إجراء الاختبار فإنه يمكن التحكم فى الخطأ β ، حيث أن قيمة (a) متممة لقيمة (β)، ولكن ذلك يرتبط بمدى خطورة أو أهمية النتائج المترتبة على الاختبار. فكلما كانت النتائج المترتبة على الوقوع فى الخطأ (α) أكثر خطورة من مثيلتها المترتبة على الوقوع فى الخطأ (a) كلما قلت قيمة (α) التى يختارها الباحث، مثل ٠,٠١ أو ٠,٠٠١، وبالتالي تزداد قيمة β ، أى أن $\beta = (١ - ٠,٠١ = ٠,٩٩)$ أو $\beta = (١ - ٠,٠٠١ = ٠,٩٩٩)$ ، أى يزداد احتمال الوقوع فى الخطأ من النوع الثانى مما يؤدي إلى نقص قوة الاختبار ($١ - \beta$) وتمثل هذه الحالة فى أغلب التجارب العملية. والعكس كلما كانت النتائج المترتبة على الوقوع فى الخطأ (β) أكثر خطورة من الوقوع فى الخطأ (α). مثل ٠,٠٥ أو ٠,١٠، وبالتالي تقل قيمة (β) أى أن $\beta = (١ - ٠,٠٥ = ٠,٩٥)$ أو $\beta = (١ - ٠,١٠ = ٠,٩٠)$ ، أى ينقص احتمال الوقوع فى الخطأ من النوع الثانى مما يؤدي إلى زيادة قوة الاختبار. وتمثل مثل هذه الحالة فى اختبار الفروض المتعلقة بأغلب المشاكل فى الجغرافية الاقتصادية.

تحديد التوزيع النظرى (الاحتمالى) للإحصائية المختبرة:

يعتمد قبول أو رفض الفروض الاحصائية، أو بمعنى آخر الاستدلال على صحة أو خطأ الفروض، على حساب بعض المقاييس الاحصائية من العينة أو العينات (والتي تعرف باحصائية الاختبار) ومقارنة هذه المقاييس بتلك المقاييس الاحصائية النظرية (أو ما يعرف بالمعايير الإحصائية) والتى عن طريقها يمكن تقدير الخطأ فى قبول أو رفض الفرض الاحصائى. فإذا كانت المقاييس الأولى تقترب من الثانية فإنه يتم قبول الفرض المختبر والعكس صحيح ويمكن اختبار القيمة المعيارية للإحصائية والتى هى، كما ذكرنا، عبارة عن الفرق بين الاحصائية المحسوبة من العينة ومعلمة المجتمع مقسوماً على الخطأ المعيارى. فمثلاً إذا كان لدينا توزيعاً عينةياً يمكن حساب المتوسط الحسابى منه ووضعه فى صورة معيارية فإن قيمة المتغير المعيارى المحسوب يمكن مقارنتها بتوزيعها النظرى، وبالتالي يمكن تقرير امكانية قبول أو رفض فرض العدم.

وتحسب قيم إحصائية الاختبار (ز) في حالة توفر بيانات عن المجتمع (أى في حالة التوزيع المعتدل)، بينما تحسب قيمة (ت)، وقيمة (ف) وقيمة مربع كاي في حالة العينات. والذي يحدد قرب قيم إحصائيات الاختبار (أى قبول أو رفض الفرض الاحصائي) هو التوزيع النظرى (الاحتمالى) لهذه الاحصائيات أى توزيع (ز) Z-Scores، توزيع (ت) T-distribution، توزيع (ف) F-distribution وتوزيع مربع كاي χ^2 -distribution. وهذه التوزيعات مدونة في جداول خاصة لرجوع إليها عند تحديد المقارنة بين الفروض النظرية والفروض الحقيقية (أنظر ملاحق الجداول الاحصائية في نهاية هذا الكتاب).

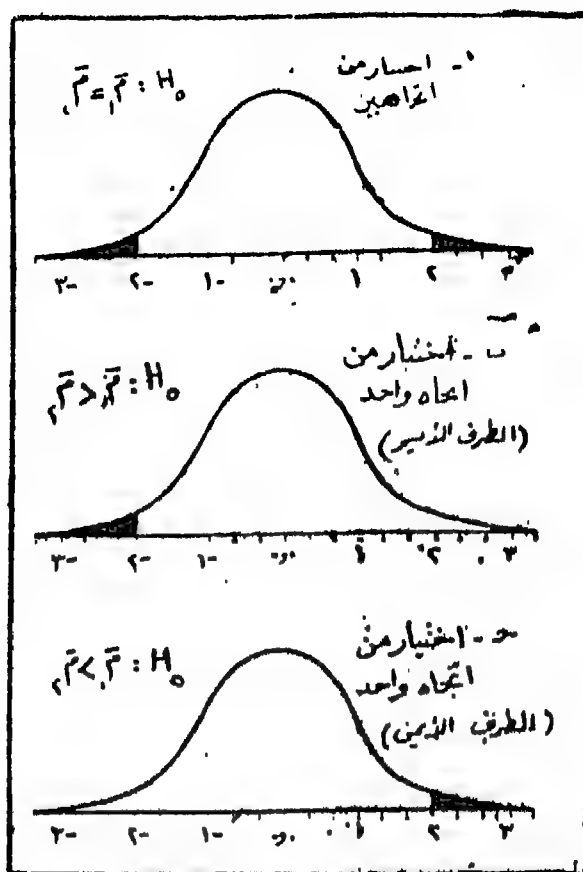
تحديد المنطقة الحرجة:

بعد تحديد قيمة مستوى للمعنوية وبعد وضع إحصائية العينة في الصورة المعيارية وتحديد التوزيع النظرى (الاحتمالى) لها يمكن تعيين منطقة الوقوع في الخطأ من النوع الأول أو ما يسمى بالمنطقة الحرجة Critical Region أو منطقة الرفض Re-jection Region، وتحتوى هذه المنطقة من منحنى التوزيع على جميع القيم الحرجة التى تدعونا إلى رفض فرض العدم وهو صحيح، وذلك لأن احتمال أن تقع نتيجة العينة في هذه المنطقة إذا كان فرض العدم صحيح يساوى مستوى معنوية (دلالة) α .

وقد تمتد المنطقة الحرجة على طرفى منحنى التوزيع النظرى (الاحتمالى) أو قد تمتد على طرف واحد من طرفى التوزيع على حسب التجربة المراد إختبارها. ويسمى الاختبار في الحالة الأولى بالاختبار ثنائى الجهة (الطرف) Two-tailed Test الذى يستخدم عندما يراد اختبار ما إذا كانت إحصائية عينة تختلف اختلافاً جوهرياً عن توقع المجتمع المفروض، وفي الحالة الثانية يسمى بالاختبار أحادى الجهة (الطرف) One-tailed Test الذى يستخدم عندما يراد اختبار الانحراف الموجب فقط أو الانحراف السالب فقط للإحصائية المحسوبة من بيانات العينة عن

متوسط المجتمع الحقيقي. فمثلاً إذا كانت الاحصائية لها توزيع معتدل وكان مستوى المعنوية $\alpha = 0,05$ وقررنا إجراء اختبار ثنائي الطرف فإن المنطقة الحرجة تشتمل في هذه الحالة على $2,5\%$ من المساحة تحت المنحنى على كل طرف من طرفيه، وتسمى المنطقة الواقعة بين المنطقة الحرجة على الطرفين بمنطقة القبول Re -gion of Acceptance وتمثل 95% من المساحة تحت المنحنى. أما في حالة الاختبار أحادى الطرف بمستوى معنوية $\alpha = 0,05$ نجد أن المنطقة الحرجة على هذا الطرف تشتمل على 5% من المساحة الكلية، فإذا كانت المنطقة الحرجة على الطرف الأيمن فإن القيم الحرجة تقع على يمين قيمة إحصائية الاختبار المعيارية، وإذا كانت المنطقة الحرجة على الطرف الأيسر فإن القيم الحرجة تقع على يسار قيمة إحصائية الاختبار المعيارية وبذلك يكون احتمال رفض فرض العدم وهو في الواقع صحيح أكبر في الاختبار أحادى الطرف منه في الاختبار ثنائي الطرف شكل (٨-٢).

ويتخذ القرار الاحصائي على أساس مقارنة النتائج المشاهدة من بيانات العينة بالنتائج النظرية، فإذا وقعت قيمة إحصائية الاختبار المحسوبة في المنطقة الحرجة يرفض فرض العدم ويقبل الفرض لبدل، ويدل ذلك على وجود فرق جوهري أو حقيقى بين الاحصائية والقيمة المفترضة لمعلمة المجتمع. أما إذا وقعت قيمة إحصائية الاختبار في منطقة القبول نعتبر الفرق بين النتيجة المشاهدة والمفروضة للمجتمع هو فرق غير جوهري أى أنه فرق ظاهري ربما يرجع إلى الصدفة المطلقة لخطأ المعاينة.



شكل رقم (٨-٢) المنطقة الحرجة واختبارات الفروض الإحصائية

وسنقوم في ما يلي بتطبيق الأسس والمبادئ لإجراء بعض اختبارات الفروض الإحصائية الخاصة بتوزيع المجتمع (التوزيع المعتدل) وذلك عن طريق حساب إحصائية الاختبار (ز).

اختبار انتماء عينة لمجتمع متوسطة معلوم:

إذا سحبنا عشوائياً عدة عينات حجمها صغير ($n < 30$) من مجتمع معتدل التوزيع متوسطة الحسابي (μ) غير معلوم، فإن التوزيع العيني يتبع توزيع المجتمع ويكون متوسطة الحسابي (\bar{x}) مناظر لمتوسط المجتمع. أما إذا سحبنا عشوائياً عدة عينات حجمها كبير ($n > 30$) من مجتمع يتصف بقرينه من الاعتدالية في

التوزيع فان متوسط التوزيع العيني (س) يقترب من التوزيع المعتدل. وعليه فإنه كلما كان حجم العينة كبيراً كلما أدى ذلك إلى اعتدالية التوزيع العيني بغض النظر عن حالة توزيع المجتمع الذى سجت منه هذه العينات.

ولاختبار الفرق بين متوسط عينة (س) ومتوسط مجتمع (م) تباينه (ع) معلوم فائنا نحسب احصائية الاختبار التى تساوى فى هذه الحالة:

$$\frac{\frac{s - m}{e}}{\frac{e}{\sqrt{n}}} = (z)$$

وفى حالة عدم معرفة تباين المجتمع (ع) فائنا نستخدم تباين العينة (ع) بدلاً منه، وفى هذه الحالة تتبع احصائية الاختبار قيم (ت) المعيارية. وباردياد حجم العينة (ن > ٣٠) فان التوزيع الأخير يقترب من التوزيع المعتدل، وتأخذ احصائية الاختبار الشكل التالى:

$$\frac{\frac{s - m}{e}}{\frac{e}{\sqrt{n}}} = (z)$$

ولتوضيح ماسبق ذكره نعطى المثال الآتى:

مثال: أراد باحث دراسة النشاط التجارى للصيدليات فى مدينة الاسكندرية مستخدماً لذلك معياراً يتمثل فى حجم مبيعاتها اليومية بالجنيه. فلما سحب الباحث عينة عشوائية مكونة من ١٤٤ صيدلية وجد أن متوسط مبيعاتها ٩٩٨,٠ جنيه، فإذا كان الانحراف المعيارى لكل الصيدليات هو ٢٠ جنيهاً فهل يعنى ذلك أن متوسط مبيعات كل الصيدليات فى الاسكندرية هو ١٠٠٠ جنيه فى اليوم وبذلك عند مستوى معنوية ٠,٠٥

لإجراء الاختبار فى المثال السابق فاننا نتبع الخطوات التالية:

١- تحديد فرض العدم والفرض البديل:

$$\bar{M} = 1000 \text{ فى مقابل } \bar{M} \neq 1000$$

٢- تحديد مستوى المعنوية أو الدلالة (أى تحديد احتمال رفض فرض العدم، أو

احتمال قبول الفرض البديل، وفى المثال قيمة $\alpha = 0.05$)

٣- تحديد إحصائية الاختبار (ز).

٤- استخدام بيانات العينة لإصدار القرار الإحصائى برفض فرض العدم أو قبول

الفرض البديل (وهو فى المثال السابق من النوع ثنائى الطرف). وباستخدام

جداول المنحنى المعتدل المعيارى نجد أن فرض العدم سيرفض إذا كانت قيمة

(ز) المحسوبة أكبر من (أو مساوية) لقيمة (ز) النظرية ± 1.96 .

وباستخدام بيانات العينة السابقة نجد أن:

$$\frac{20}{\sqrt{144}} = \frac{\text{الانحراف المعيارى}}{n} = \text{الخطأ المعيارى}$$

$$1.66 = \frac{20}{12} =$$

وبحساب قيمة (ز) نجد أن:

$$1.204 = \frac{2 - 1000 - 998.0}{1.66} = \frac{\bar{M} - \bar{M}_0}{\frac{s}{\sqrt{n}}} = (z)$$

وبمقارنة قيمة (ز) المحسوبة وهى -1.204 فى المثال بالقيمة المستخرجة من

جداول المنحنى المعتدل المعيارى عند مستوى معنوية 0.05 وهى -1.96 نجد أن

قيمة (ز) المحسوبة لاتقع فى منطقة الرفض، وبناء على ذلك نقبل فرض العدم

القائل بأن متوسط مبيعات الصيدليات فى الاسكندرية ١٠٠٠ جنيه فى اليوم الواحد.

مثال (٢):

فى المثال السابق إذا كان متوسط مبيعات الصيدليات فى العينة هو ٩٩٥ جنيه فى اليوم والمطلوب اختبار الفرض القائل بأن هذا المتوسط أكبر من أو يساوى المتوسط العام لجميع الصيدليات ١٠٠٠ جنيه فى اليوم فى مقابل الفرض البديل بأن المتوسط فى العينة أقل من ١٠٠٠ جنيه وذلك عند مستوى معنوية ٠,٠٥. باتباع نفس الخطوات السابقة يلاحظ أن:

$$H_0: \mu \leq 1000 \quad H_1: \mu > 1000$$

ويلاحظ أن الخطأ فى الفرض البديل هو خطأ من النوع الثانى β لاختبار أحادى الطرف. أى أن فرض العدم سيرفض عندما تكون μ أصغر من ١٠٠٠ جنيه بدرجة المعنوية المذكورة. ومنطقة الرفض ستقع بائالى على الطرف الأيسر للمنحنى المعتدل المعيارى شكل (٨-٢ ب) وبحساب قيمة (ز) نجد أن:

$$z = \frac{995 - 1000}{1.66} = -3.01$$

وبمقارنة قيمة (ز) المحسوبة من العينة (-٣,٠١) بالقيمة المستخرجة من الجدول وهى ١,٦٤٠ نجد قيمة (ز) المحسوبة أقل من قيمة (ز) بادل ، أى أنها تقع فى منطقة الرفض. وبذلك نستطيع رفض فرض العدم للقائل بأن متوسط مبيعات الصيدليات فى العينة أكبر من أو يساوى المتوسط العام وهو ١٠٠٠ جنيه وذلك عند مستوى معنوية ٠,٠٥ أو بمعنى آخر قبول الفرض البديل القائل بأن متوسط مبيعات الصيدليات فى العينة أقل من ١٠٠٠ جنيه عند نفس مستوى المعنوية أو الدلالة.

مثال (٣):

فى المثال الأول إذا كان متوسط مبيعات الصيدليات هو ١٠٥ جنيه فى اليوم الواحد والمطلوب اختبار الفرض القائل أن متوسطات المبيعات فى العينة أقل أو يساوى ١٠٠٠ جنيه فى مقابل الفرض البديل بأن متوسط المبيعات للعينة أكبر من ١٠٠٠ جنيه فى اليوم عند مستوى دلالة أو معنوية ٠,٠٥ فى هذا المثال ينحصر الاختبار فى:

$$H_0: \bar{X} \geq 1000 \quad \text{فى مقابل} \quad H_1: \bar{X} < 1000$$

وفى هذا المثال نجد الفرض لا بديل عكس نفس الفرض فى المثال الثانى أو بمعنى آخر نجد أن الفرض البديل ذو طرف أيمن أى أنه يمكن رفض فرض العدم عندما تكون \bar{X} أكبر من ١٠٠٠ جنيه بدرجة معنوية ٠,٠٥ (شكل ٨ - ٢ ج).
ويحساب قيمة (ز) نجد أن:

$$Z = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} = \frac{1000 - 1005}{\frac{1,66}{\sqrt{30}}} = -3,01$$

وبمقارنة قيمة (ز) المحسوبة وهى -٣,٠١، بالقيمة النظرية فى جداول المنحنى المعتدل المعيارى عند مستوى معنوية ٠,٠٥ وهى -١,٦٤٥، نجد أن قيمة (ز) المحسوبة أكبر من ١,٦٤ أى قيمة (ز) المحسوبة من العينة تقع فى منطقة الرفض، لذلك نرفض فرض العدم القائل بأن متوسط المبيعات فى العينة أقل من المتوسط العام للمبيعات لكل الصيدليات فى الاسكندرية عند مستوى معنوية ٠,٠٥، وقبول الفرض البديل الذى يقول أن متوسط العينة أكبر من المتوسط العام عند نفس مستوى الدلالة أو المعنوية.

اختيار الاختبارات الإحصائية:

يعتمد إختيار الاختبارات الإحصائية Statistical Tests على طبيعة وخصائص البيانات Characteristics of the data، والقيمة الفعلية للأساليب الكمية المرتبطة بها والمستخدمة في عمليات البحث والتحليل، والافتراضات الإحصائية عن المجتمعات التي تستقى منها البيانات. وكما ذكرنا في مقدمة هذا الكتاب أن هناك ثلاثة أنواع من البيانات، حسب الطرق المختلفة التي تقاس بواسطتها، وهى البيانات الرسمية (النوعية) أو الوصفية Nominal data والبيانات الترتيبية Ordinal or Rank ing data وبيانات الفترة Interval data. وكما عرفنا أن النوع الأول من البيانات يتصف بأنه قائم على أساس التعداد أو العد Counts بينما تكون القياسات Mes- surements أهم صفات النوعين الآخرين. كذلك قد تكون البيانات ذات قيم فردية أو ثنائية (أى مزدوجة) على أساس أن العد أو القياس فى مجموعة بيانات يماثل نظيرة فى المجموعات الأخرى - بالإضافة إلى أننا قد نكون بصدد مقارنة بيانات فعلية (حقيقية) لمجموعة واحدة، أو لمجموعتين أو أكثر بيانات توزيع نظرى. لكل ذلك فإن أنواع الاختبارات الإحصائية المستخدمة فى البحوث الجغرافية تختلف حسب نوعية البيانات والطرق التى قياست بها. فهناك إختبارات إحصائية لا تصلح أو لا يمكن تطبيقها إلا فى حالات بيانات الفترة بل أن معظم الأساليب تفترض أن البيانات المتوفرة هى من هذا النوع، أما البيانات الإسمية (النوعية) والترتيبية فلا يستخدم معهما إلا الأساليب الإحصائية البسيطة. ويوضح الجدول التالى (جدول رقم: ٨-٢) بعضا من الشروط التى يجب أن يلم بها الباحث قبل إختيار، كما يبين الأنواع المختلفة من الاختبارات الإحصائية وملائمة كل منها لخصائص وطبيعة البيانات.

جدول رقم (٨-٧)
خصائص البيانات وأنواع الاختبارات الإحصائية
المستخدمة في مقارنة الاختلاف بين قيم التوزيعات

بيانات خام	نوع الجدولة	نوع الاختبار
مقارنة مجموعتين من البيانات	١- بيانات قائمة على العد ينتج عنها تكرارات إسمية: أ- في فئتين ب- في أكثر من فئتين	قيم عددية قيم عددية مربع كاي مربع كاي
	٢- قياسات فردية من نوع بيانات الفترة	تكرارات رتب قيم عددية مربع كاي أو كولموجوروف سميرنوف مان - هوتيني (اختبار «ي») ستيوونت (اختبار «ت»)
	٣- قياسات ثنائية من نوع بيانات الفترة	رتب قيم عددية ويلكوكسون ستيوونت (اختبار «ت»)
مقارنة أكثر من مجموعتين من البيانات	١- بيانات قائمة على العد ينتج عنها تكرارات إسمية	قيم عددية مربع كاي
	٢- قياسات من نوع بيانات الفترة (فردية وثنائية غير مستكافئة في عدد مفرداته)	تكرارات رتب قيم عددية مربع كاي كروسكال - واليس (اختبار «ه») تحليل التباين (اختبار «ف»)

أما من حيث القيمة الفعلية التي يتوقف عليها اختيار الاختبارات الإحصائية فنقصد بها القوة Power أو القدرة على التمييز بين الفرض الحقيقي والفرض غير الحقيقي، ويعتمد ذلك على حجم لعينات المختبرة من ناحية وعلى مدى فاعلية الاختبار نفسه من ناحية أخرى. فمثلاً إذا اخترنا أسلوباً بسيطاً في البحث والتحليل فإن ذلك يتطلب سحب عينات كبيرة الحجم حتى يتحقق نفس مستوى القوة أو التمييز الذي تتصف به الأساليب ذات المقدرة والفاعلية. وعليه فإننا نتوقع أن تكون هناك مقاييساً أكثر قوة وفاعلية يمكن استخدامها في التحليل الإحصائي أكثر من غيرها حتى إذا كانت العينات المختبرة صغيرة الحجم. فمثلاً إذا كان حجم العينة المطلوب لتحقيق قوة أحد الاختبارات ذات الفاعلية في التحليل هو n_1 ، وكان حجم العينة لتحقيق مستوى نفس القوة لاختبار آخر أقل فاعلية هو n_2 ، فإن قوة الاختبار الأخير تساوي $(n_1 \div n_2) \times 100\%$ ، وتبعاً لذلك فإن اختبار له قوة تساوي 80% ($\frac{4}{5}$) سوف يتطلب حجم عينة مقدارة 125% ($\frac{5}{4}$) من حجم العينة التي يتطلبها أكبر الاختبارات المتاحة قوة وأكثرها فاعلية.

أما عن الافتراضات الإحصائية عن المجتمعات ونوع توزيعاتها التكرارية فتبدو أهميتها في أنها تتخذ كأساس لتقسيم «مجتمع» الاختبارات الإحصائية المستخدمة في المقارنة، إلى «أُسَرتين» هما: (١) الاختبارات الكلاسيكية (القديمة) Classical أو البارامترية (المعلمية) Parametric tests، التي سادت كأسلوب عمل في النظرية الإحصائية والممارسة العملية حتى وقت ليس بعيد، والتي يمكن تطبيقها في حالات بيانات الفترة التي هي أكثر شيوعاً لأنها أكثر توافقاً مع افتراضات هذه الاختبارات عن المجتمعات. (٢) الاختبارات الحديثة أو اللابارامترية (غير اللامعلمية) Non-Parametric التي اكتسبت أهمية خاصة منذ الحرب العالمية الثانية ٣٩ - ١٩٤٥ إذ أنه يمكن تطبيقها على البيانات الإسمية (النوعية) والترتيبية وبيانات الفترة على حد سواء. وتعد اختبارات النوع الأول أكثر قوة من اختبارات النوع الثاني، ولأنها نفترض أن مفردات المجتمع تتوزع توزيعاً معتدلاً كما أنها تستلزم في حالة استخدام عينات صغيرة الحجم في التحليل أن يكون التوزيع

المعتدل للمجتمع الذى سحبت منه هذه العينات مؤكداً، بينما لا تتقيد بذلك الاختبارات اللاباراميتريّة.

وفى الحالات التى لا يكون فيها توزيع بيانات المجتمع معتدلاً يتم تحويل البيانات بطرق مختلفة، سبق شرحها، ليصبح توزيعها معتدلاً، حتى يمكن تطبيق الاختبارات الباراميتريّة عليها، فإن لم يتحقق ذلك فيجب تطبيق الاختبارات اللاباراميتريّة لأنها لا تشترط توزيعاً معتدلاً للبيانات. كما تبرز أهمية مثل هذه الاختبارات فى حالة إذا كانت العينات قيد الاختبار صغيرة الحجم، وهو ما سنتناقشة بالتفصيل فى الفصلين التاليين.

الفصل التاسع
أساليب المقارنة الباراميتريّة
(المعلمية)

الفصل التاسع

أساليب المقارنة البارامترية (المعلمية)

تتطلب الاختبارات البارامترية Parametric Tests والأساليب الكمية المرتبطة بها - لمقارنة معالم المجتمعات أو إحصائيات العينات - توفر الخصائص التالية في بيانات المجتمع Population قيد الفحص:

- ١- أن يكون توزيع البيانات توزيعاً معتدلاً (متماثلاً)، أى أن معامل التواءه يساوى صفراً.

- ٢- أن تكون المفردات المشاهدة أو الحالات (Observations or Cases) مستقلة عن بعضها البعض، أى أن اختيار إحدى المفردات لا يمنع إمكانية اختيار أى مفردة أخرى من المفردات المطلوب دراستها.

- ٣- أن يكون للمجتمعات المقارنة مع بعضها البعض تبايناً Variance متساوياً، أو بمعنى آخر أن يكون هناك تجانساً بين المجتمعات موضع المقارنة.

- ٤- أن تكون البيانات المقاسة والتي تجرى عليها الاختبارات من نوع بيانات الفترة Interval Data .

فإذا لم تتوفر هذه الشروط أو الخصائص في البيانات، فإن تطبيق الاختبارات البارامترية عليها يكون غير مناسب وبالتالي تكون النتائج النهائية مضللة، ولهذا نلجأ إلى النوع الآخر من الاختبارات: لاختبارات غير البارامترية - وهى الاختبارات التى سنعرضها وسناقش الأساليب المرتبطة بها فى الفصل التالى مباشرة (الفصل العاشر).

وتتم عملية مقارنة وتحليل البيانات المعتدلة التوزيع، لاختبار الفروق بين معالم

المجتمع وبين احصائية (المتوسطات الحسابي أو الانحراف المعياري) عينتين أو أكثر
لمتغير واحد، ومعايرة نتائجها بواسطة عدة اختبارات باراميتريّة أكثرها شيوعاً هو
اختبار ستودنت (ت) Student-t test واختبار تحليل التباين (أو نسبة «ف»)
Analysis of Variance (F ratio test) وسنناقش في هذا الفصل كل
اختبار على حدة من حيث أهميته وطرق حسابه وأهم مجالات ومشاكل تطبيقه.

اختبار ستودنت - (ت) (١)

(اختبار الفرق بين المتوسطات)

أوضحنا في الفصل السابق أن الاختبارات الاحصائية الخاصة بالعينات تفترض
في أغلب الأحيان أن تكون بيانات العينة موزعة توزيعاً معتدلاً. وكما عرفنا أن
توزيع العينة قد لا يكون كذلك، لذا فإن من الواجب إجراء بعض التعديلات في
البيانات ليتسنى لها الاقتراب من التوزيع المعتدل. وذكرنا أيضاً أن بيانات العينة
مهما كانت متماثلة فإنها لن تعكس تماماً خصائص المجتمع الذي سحبت منه،
وبالتالي توجد بعض الفروق بين خصائص العينة ومعاليم المجتمع الذي تمثله والتي
يمكن تقديرها على أساس احتمالات أخطاء معينة (مستويات المعنوية α أو
مستويات الثقة β).

وكما ذكرنا لا نستطيع دائماً قياس جميع المفردات في المجتمع لمعرفة معلّمته
(المتوسط الحسابي)، ولكن يستعاض عن متوسط المجتمع بمتوسط عينة حجمها
كبير. وحتى يكون متوسط العينة ممثلاً ومماثلاً لمتوسط المجتمع يجب أن يكون الفرق
بين المتوسطين صغيراً. أي إذا كان هناك فرض يقول أن متوسط العينة يساوي
متوسط المجتمع فأننا نقبل هذا الفرض وذلك على العكس إذا كان الفرق بين
المتوسطين كبيراً فأننا نرفض الفرض السابق حيث أنه في هذه الحالة سيكون هناك

(١) اكتشف العالم البريطاني William S. Gosset التوزيع الإحتمالي «ستودنت - ت» في
سنة ١٩٠٨ ولم يشأ أن يذكر اسمه فنشره بأضياء ستودنت (أي طالب Student) كبديل مستعار
لإسمه، وأعطى الحرف الأخير في الكلمة وهو (ت = t) كأسم للاختبار الذي يستخدم هذا التوزيع في
المعايرة الإحصائية.

اختلاف جوهري بين متوسط العينة ومتوسط المجتمع. وتعني كلمة الاختلاف الجوهري Significant إحصائياً أن معلمة المجتمع تختلف اختلافاً كبيراً ولا تتفق مع نتائج العينة المسحوبة.

وبالمثل عند إجراء البحوث والدراسات الجغرافية تقابلنا كثيراً من المشاكل التي تتطلب المقارنة بين متوسطي عينتين لمعرفة ما إذا كانت هاتان العيتان مسحوبتين من مجتمعين مختلفين ولهما نفس المتوسط أم مسحوبتين من نفس المجتمع. فمثلاً قد نجد من الأفضل عملياً عند اختبار مدى فاعلية عامل معين أن نسحب عينتين الأولى لتمثل المجتمع قبل تأثير هذا العامل والثانية لتمثل المجتمع بعد تأثير العامل ونختبر ما إذا كان الفرق بين متوسط العينتين فرق جوهري أو غير جوهري، فإذا كان الفرق جوهري نستنتج فاعلية العامل. أما إذا كان الفرق غير جوهري فإننا نستنتج عدم فاعلية هذا العامل. وأن الفرق قد يكون راجعاً للمصدقة المطلقة أو قد يكون ناتجاً من خطأ المعاينة. ولتوضيح ذلك نقول أنه عند استخدام نوعين مختلفين من المخصبات الزراعية لمعرفة ما إذا كان لهما تأثيراً واضحاً على نوع معين من التربة في منطقتين لهما نفس الظروف، وبالتالي على الإنتاج الزراعي أم لا، تسحب عينة من المجتمع الأول وعينة من المجتمع الثاني ثم يحسب المتوسط الحسابي لكل عينة على حدة، ثم يجرى الاختبار على قيم المتوسطتين. فإذا كانت \bar{M}_1 ، \bar{M}_2 هما متوسطي المجتمعين الأول والثاني على التوالي، وكانت S_1 ، S_2 هما متوسطي العينتين المسحوبتين من المجتمعين السابقين فإن الفرق ($S_1 - S_2$) هو متغير عشوائي الفرق ($M_1 - M_2$). ويحدد الفرق بين المتوسطتين بواسطة الوحدات المعيارية، أي بتحويل القيم الأصلية إلى وحدات معيارية ومعايرتها بوحدات معيارية بوحدات معيارية نظرية حتى نستطيع الحكم على أن هذا الفرق هو فرق جوهري أم لا. وهناك مجالات أخرى مشابهة منها على سبيل المثال لا الحصر: اختبار وجود فروق جوهرية بين مستوى كفاءة عمال الإنتاج في مصنعين مختلفين، أو اختبار وجود فروق جوهرية بين درجة صلابة نوعين من الصخور. فإذا ثبت أن الفرق بين كل عيتين هو فرق جوهري أو حقيقى فإن ذلك يكون دليلاً على أن العيتين مسحوبتان من مجتمعين مختلفين، أما إذا ثبت أن الفرق غير جوهري فإن ذلك

يعنى أن الفرق بين متوسطى العينتين يرجع الخطأ الصدفة أو لخطأ المعينة وأن العينتين قد تكونا مسحورتان من مجتمع واحد أو من مجتمعين لهما نفس المتوسط الحسابى.

ويستخدم أسلوب أو اختبار ستيودنت (ت) لاختبار المتوسطات فى حالة إذا لم يكن تباين المجتمع معلوماً والذي يستبدل بتباين العينات. وبما أن تباين أية عينة لا يساوى بالضبط تباين المجتمع المسحوبة منه، لذلك فإننا لو استخدمنا توزيع «ز» فإن اختبار المتوسطات الحسابية سيتعرض للخطأ. غير أنه من الملاحظ على الاختبارين «ز»، «ت» أن قيم «ت» النظرية تصبح تقريباً نفس قيم توزيع «ز» إذا زاد حجم العينة عن ٩٠ مفردة. لذلك ففى الحالات التى يكون فيها حجم العينة أكبر من ٩٠ مفردة. لذلك ففى الحالات التى يكون فيها حجم العينة أكبر من ٩٠ مفردة فإنه يمكننا استخدام إما توزيع «ت» أو توزيع «ز» للاستدلال على صحة الفرض الموضوع لاختبار المتوسطات. كما يلاحظ أن قيمة «ت» تقاس أيضاً بوحدات الخطأ المعيارى للمجتمع المقدر من بيانات العينة ولذلك فإن توزيع «ت» يمثل توزيعات للمتوسطات الحسابية للعينات، ولكن نظر لأن توزيع قيم «ت» يعتمد على حجم العينة أو بالأحرى على درجات الحرية فإنه يلزم عمل توزيع احتمالى لكل درجة من درجات الحرية والجدول المختصر فى ملاحق الكتاب الذى يبين قيمة «ت» عند مستويات معنوية مختلفة يعتبر كافياً لاستخدامه فى العينات الصغيرة. وتحسب درجات الحرية لعينة واحدة عد مفرداتها «ن» بطرح مفردة واحدة من هذه المفردات (أى ن - ١)، ويساعد ذلك فى تحديد تباين العينة والذي هو عبارة عن مجموع مربعات انحرافات القيم عن متوسطها الحسابى مقسوماً على درجات الحرية للعينة. ويمكن وضع ذلك فى الصيغة الرياضية الآتية:

$$\sigma^2 = \frac{\text{مجم (س - س')^2}}{n - 1}$$

حيث σ^2 هى تباين العينة، (س - س') هى انحرافات القيم عن متوسطها الحسابى.

وهناك عدة شروط يجب توافرها عند استخدام اختبار «ت» وبغيرها فان النتائج التي نتوصل إليها لا تكون صحيحة:

أولاً: استقلال مفردات العينات قيد الاختبار أى أن كل عينة تكون مسحوبة بطريقة عشوائية ومستقلة عن الأخرى.

ثانياً: أن يكون التوزيع التكرارى للصفة المتغيرة لكل عينة توزيعاً معتدلاً.

ثالثاً: أن يكون هناك تجانس بين العينات، ويقصد هنا بالتجانس مدى التفاوت بين تباين أى عيتين، ويقاس هذا المدى بقسمة التباين الأكبر على التباين الأصغر، وليس بإيجاد الفرق بين تباين العيتين.

رابعاً: يجب لا يكون الفرق بين متوسطى العيتين كبيراً، وذلك لأن حجم العينة يؤثر على مستوى معنوية أو دلالة قيمة «ت»، وأن هذا المستوى يتأثر إلى حد كبير بدرجات الحرية الذى بدوره يحدد المنطقة الحرجة أو منطقة رفض فرض العدم الخاص باختبار.

اختبار المتوسطات للعينات الكبيرة (ن > ٣٠):

عند اختبار عيتين عشوائيتين (مستقلتين) من مجتمعين مختلفين وكان حجم العينة الأولى هو n_1 ومتوسطها هو \bar{x}_1 وتباين مجتمعها هو σ_1^2 وتباين المتوسط $= \frac{\sigma_1^2}{n_1}$ ، بينما كان حجم العينة الثانية هو n_2 وتباين مجتمعها هو σ_2^2 وتباين المتوسط $= \frac{\sigma_2^2}{n_2}$ ، فإن تباين الفرق بين المتوسطين يمكن حسابه من المعادلة الآتية:

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = \sigma_1^2 - \sigma_2^2 + \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \quad (٩-١)$$

ويكون الانحراف المعياري (الخطأ المعياري) للفرق بين المتوسطين عبارة عن الجذر التربيعي للتباين أى أن:

$$ع س - س_2 = \sqrt{\frac{\frac{2}{2} + \frac{2}{1}}{2}} \dots\dots\dots (2-9)$$

ويتم بذلك حساب مقياس (ت) من المعادلة الآتية:

$$ت = \frac{\text{الفرق بين متوسطى العينتين}}{\text{الخطأ المعيارى لهذا الفرق}}$$

$$ت = \dots\dots\dots (3-9) \sqrt{\frac{\frac{2}{2} + \frac{2}{1}}{2}}$$

وتتوزع القيمة اتى نحصل عليها من هذا المقياس تبعاً (ت) بدرجات الحرية $(2 - 2 + 1)$

وإذا كان حجم كل من العينتين n_1, n_2 كبيراً وأن كلا منهما مسحوبة من مجتمع مختلف وحسبنا المتوسط الحسابى لكل عينة وكان الفرق بينهما هو $(\bar{s}_1 - \bar{s}_2) = A$ ، وإذا سحبنا الفرق بين متوسطى العينتين ثم وضعنا هذه الفروق فى شكل توزيع تكرارى نجد أن متوسط هذه لفروق $(\frac{A}{n})$ حيث n هى عدد العينات) يتوزع توزيعاً قريباً من التوزيع المعتدل. وإذا كان تباين المجتمعين متساويين $(\sigma_1^2 = \sigma_2^2)$ فإنه يمكن حساب التباين الكلى للمجتمع ثم الصيغة الإحصائية الآتية:

$$\sigma^2 = \frac{مج (س_1 - س_1) + 2(س_2 - س_2)}{2 - 2 + 1}$$

أما إذا تساوت المفردات فى العينتين (أى أن $n_1 = n_2 = n$) فلإن الخطأ المعيارى للفرق بين المتوسطين سيكون فى هذه الحالة عبارة عن:

$$ع س_1 - س_2 = \sqrt{\frac{\sigma^2}{n}}$$

حيث E^2 هي التباين المشترك بين العينتين والذي يساوى:

$$E^2 = \frac{\text{مجموع مربعات الانحرافات للعينتين}}{\text{مجموع درجات الحرية}}$$

وبذلك يمكن حساب الخطأ المعياري للفرق بين المتوسطين بالصيغة الاحصائية الآتية:

$$E^2 = \sqrt{\frac{2}{n} \left[\frac{\text{مجموع} (S_1 - \bar{S})^2 + \text{مجموع} (S_2 - \bar{S})^2}{2 - n_1 + n_2} \right]}$$

وتكون قيمة (ت) المحسوبة فى هذه الحالة هي:

$$t = \sqrt{\frac{2}{n} \left[\frac{\text{مجموع} (S_1 - \bar{S})^2 + \text{مجموع} (S_2 - \bar{S})^2}{2 - n_1 + n_2} \right]} \dots (9-4)$$

وبمقارنة قيمة (ت) المحسوبة لكل من الحالتين والقيمة المناظرة لها فى جداول توزيع (ت) بمدلول درجات الحرية وعند مستوى معنوية معين يمكن رفض فرض العدم إذا كانت قيمة (ت) المحسوبة أكبر من قيمة (ت) النظرية. والعكس يكون صحيحاً (أى نقبل فرض العدم) إذا كانت قيمة (ت) المحسوبة أقل من قيمتها النظرية عند مستوى الدلالة (المعنوية) المعين.

مثال (١):

أخذت عينتين من إنتاج حقول الفحم فى منطقة ما لفترة عشرة سنوات، فكانت بياناتهما كمايلي:

العينة الأولى	العينة الثانية	
١٠٠	١٠٠	حجم العينة
٣٥	٢٤	الوسط الحسابى
٤	٥	الانحراف المعيارى

والمطلوب اختيار الفرض القائل بأن متوسطى المجتمعين متساويات فى مقابل الفرض البديل القائل أن متوسط إنتاج حقول مجتمع العينة الأولى أكبر من متوسط مجتمع حقول العينة الثانية وذلك بمستوى معنوية ٠,٠٥ .

$$\text{الفرض: } H_0: \bar{y}_M = \bar{y}_m \text{ فى مقابل } H_1: \bar{y}_M < \bar{y}_m$$

$$\text{أو } H_0: (\bar{y}_M - \bar{y}_m) = \text{صفر فى مقابل } H_1: (\bar{y}_M - \bar{y}_m) \neq \text{صفر}$$

وبحساب قيمة (ت) من العينة بافتراض أن $\bar{y}_M - \bar{y}_m = \text{صفر}$ نجد أن:

$$t = \frac{\frac{1}{\frac{25}{100} + \frac{16}{100}}}{\frac{34 - 35}{\frac{2(5)}{100} + \frac{2(4)}{100}}} =$$

$$1,06 = \frac{1}{0,64} = \frac{1}{0,41} =$$

وحيث أن قيمة (ت) المحسوبة من العينة (١,٥٦) أقل من القيمة (ت) = ٢,١ المناظرة لدرجات الحرية (ن_١ + ن_٢ - ٢ = ١٠ + ١٠ - ٢ = ١٨) فى جدول توزيع (ت) بمستوى معنوية ٠,٠٥ فإن القيمة المحسوبة لاتقع فى منطقة رفض الفرض، وعلى ذلك فإن البيانات الخاصة بالعينتين قيد الاستقصاء غير كافية لرفض فرض العدم عند مستوى معنوية أو دلالة ٠,٠٥ ونستنتج من ذلك أن الفرق (س_١ - س_٢) هو فرق يرجع للصدفة المطلقة أو لأخطاء المعاينة وليس فرق جوهرياً بمستوى معنوية (α) = ٠,٠٥ .

اختبار المتوسطات للعينات الصغيرة (ن > ٣٠):

ذكرنا أنه فى حالة إذا كان الانحراف المعياري للمجتمع غير معلوم وكانت العينة كبيرة الحجم (ن > ٣٠) فإنه يمكن حساب الخطأ المعياري للمتوسط الحسابي للمجتمع من واقع الانحراف المعياري المحسوب للعينة (ع) بدلاً من

الانحراف المعياري المجهول (ع) للمجتمع. ولكن قد نضطر إلى سحب عينة صغيرة (ن > ٣٠) ومنها يمكن الحصول على تقدير غير متحيز للمعلمة (ع)، وذلك باستخراج الجذر التربيعي لخارج قسمة مجموع مربعات الانحرافات عن المتوسط الحسابي للعينة على عدد المتغيرات المستقلة الخطية (درجة الحرية ن - ١)، لأن العينة الصغيرة تعطي معلومات عن المجتمع أقل دقة من بيانات العينات الكبيرة. وفي مثل هذه الحالة يتبع نفس الأسلوب الاحصائي السابق، أى يبنى الاستنتاج الاحصائي على أساس التوزيع الاحتمالي (ت)، فيصمم الاختبار الاحصائي لمقارنة الفرق بين متوسطي عيّنتين صغيرتين وذلك في ضوء تقدير الخطأ المعياري للفرق بين المتوسطين من واقع البيانات المشاهدة للعينتين المستقلتين بافتراض أنهما مسحوبتان من نفس المجتمع وفي هذه الحالة نستخدم الصيغة الآتية للقيمة المعيارية (ت):

$$T = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{s_2^2}{n_2} + \frac{s_1^2}{n_1}}} \quad \text{..... (٩ - ٥)}$$

وهذه القيمة المعيارية لها توزيع احتمالي (ت) بدرجات الحرية (ن_١ + ن_٢ - ٢) وبما أن قيمة الانحراف المعياري للمجتمع (ع) غير معلومة، فإنه يمكن الحصول على تقدير غير متحيز لهذه القيمة من واقع عيّنتين مستقلتين كما يلي:

$$\bar{x}_2 - \bar{x}_1 = \frac{\text{محد (س - س)}_2}{n_2} + \frac{\text{محد (س - س)}_1}{n_1} \quad \text{..... (٩ - ٦)}$$

$$\frac{\text{محد (س - س)}_2}{n_2} = \text{لكن } s_2^2$$

$$\therefore \text{محد (س - س)}_2 = n_2 s_2^2$$

وبالتعويض عن محد (س - س)_٢ في المعادلة (٩ - ٦) نجد أن:

$$C = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{n_2 s_2^2}{n_2} + \frac{n_1 s_1^2}{n_1}}} \quad \text{..... (٩ - ٧)}$$

ولكن الخطأ المعياري للعينة الأولى $= \sqrt{\frac{E}{n_1}}$ ، والخطأ المعياري للعينة الثانية $= \sqrt{\frac{E}{n_2}}$ ، والخطأ المعياري للفرق بين متوسطي عينتين مسحوبتين من نفس المجتمع هو :

$$ع س_1 - ع س_2 = \sqrt{\frac{E}{n_1} + \frac{E}{n_2}} \times ع \quad \dots (9-8)$$

وبالتعويض عن قيمة (ع) في المعادلة (9-7) نجد أن:

$$ع س_1 - ع س_2 = \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \times \sqrt{\frac{n_1 E_1 + n_2 E_2}{n_1 + n_2 - 1}} \quad \dots (9-9)$$

وبذلك يمكن وضع الاختيار (ت) على الصورة المعيارية الآتية :

$$ت = \frac{ع س_1 - ع س_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \times \sqrt{\frac{n_1 E_1 + n_2 E_2}{n_1 + n_2 - 1}}} = \frac{\text{الفرق بين متوسطي العينتين}}{\text{الخطأ المعياري للفرق}}$$

..... (9-10)

وفي حالة تساوى عدد المفردات لكل عينة من العينتين (أى أن $n_1 = n_2 = n$) فإن الصورة المعيارية للاختبار تكون على النحو التالي :

$$ت = \frac{ع س_1 - ع س_2}{\sqrt{\frac{E_1 + E_2}{1 - n}}} \quad \dots (9-11)$$

مثال (٢) :

فى تجربة لمعرفة تأثير نوعين من السماد على محصول القطن سمدت ١٠ أفدنة من السماد الأول فكان متوسط إنتاجها ٧,٥ قنطاراً للفدان الواحد بانحراف

معياري قدره ٠,٦ قنطاراً، كما سمحت ٩ أفدنة أخرى من نفس درجة خصوبة التربة بالسماح من النوع الثاني فانتج في المتوسط ٦,٨ قنطاراً للفدان الواحد بانحراف معياري ٠,٥ قنطاراً، فهل يمكن الحكم بأن النوع الأول من السماح أفضل من النوع الثاني وذلك بمستوى معنوية ٠,٠٥ .

$$١ - \text{الفرض } H_0 : \bar{S}_1 \geq \bar{S}_2$$

$$, \text{الفرض } H_1 : \bar{S}_1 < \bar{S}_2$$

$$٢ - \text{مستوى المعنوية} = ٠,٠٥$$

$$٣ - \text{القيمة المعيارية للاختبار هي:}$$

$$T = \frac{\bar{S}_2 - \bar{S}_1}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \times \sqrt{\frac{2s_{e2}^2 + 1s_{e1}^2}{2 - n_2 + 1n_1}}$$

$$٤ - \text{توزيع ت الاحتمالي (النظري) بدرجات الحرية (ن + ١ - ٢ن - ٢) = (١٠ + ٢ - ٩ = ١٧)}$$

٥ - منطقة الرفض: طبقاً للقواعد السابقة فإنه يمكن رفض فرض العدم إذا كانت قيمة (ت) المحسوبة من البيانات المشاهدة أكبر من ٢,١١ في اختبار الطرفين، ولكن في اختبار الطرف الواحد نرفض فرض العدم إذا كانت (ت) أقل من -١,٧٤ .

$$٦ - \text{تحسب قيمة (ت) من البيانات المشاهدة للعينتين كمايلي:}$$

$$\begin{array}{ll} \bar{S}_1 = ٧,٥ & \bar{S}_2 = ٦,٨ \\ \bar{e}_1 = ٠,٥ & \bar{e}_2 = ٠,٦ \\ n_1 = ١٠ & n_2 = ٩ \end{array}$$

$$T = \frac{٦,٨ - ٧,٥}{\sqrt{\frac{1}{10} + \frac{1}{9}}} \times \sqrt{\frac{2(٠,٥)^2 + 1(٠,٦)^2}{2 - ٩ + ١٠}}$$

$$4,516 = \frac{,7}{,100} =$$

٧- الاستنتاج: بما أن قيمة (ت) المحسوبة (٤,٥١٦) أكبر من قيمة (ت) النظرية (٢,١١) فمعنى ذلك أنها تقع ضمن منطقة الرفض (المنطقة الحرجة)، وبهذا نرفض فرض العدم ونقبل الفرض البديل أى أن هناك فرق جوهري بين النوعين من السماد فى درجة تخصيبهما للتربة، ونستنتج أن النوع الأول من السماد ينتج محصولاً أوفر من النوع الثانى من السماد.
مثال (٣):

إذا كانت لدينا عيتتين مستقلتين من عمال مصنعى أُسمنت وكان عدد عمال العينة الأولى ٩ عاملاً وعدد عمال العينة الثانية ١٥ عاملاً وكان متوسط إنتاج عمال العينة الأولى ٥٠ طناً فى الشهر بانحراف معيارى ٧,٥ طن، ومتوسط إنتاج عمال العينة الثانية ٥٥ طناً فى الشهر وبانحراف معيارى ٦ طن، فهل يمكن الحكم على أن متوسط الإنتاج الشهرى فى المصنعين متساوى؟

$$١- \text{الفرض } H_0: \mu_1 = \mu_2, \text{ الفرض } H_1: \mu_1 \neq \mu_2$$

$$٢- \text{مستوى المعنوية} = ٠,٠٥$$

$$٣- \text{درجات الحرية للتوزيع الإحتمالى (ت)} = ٩ + ١٥ - ٢ = ٢٥$$

٤- منطقة الرفض: تبعاً للفروض السابقة نجد أن قيمة (ت) النظرية التى تحدد منطقة الرفض هى $\pm ٢,٠٦$ أى تقبل فرض العدم إذا كانت (ت) أقل من + ٢,٠٦ أو أكبر من - ٢,٠٦.

٥- حساب قيمة (ت) من البيانات المشاهدة على النحو التالى:

$$\mu_2 = ٥٥$$

$$\mu_1 = ٥٠$$

$$\sigma_2 = ٦$$

$$\sigma_1 = ٧,٥$$

$$n_2 = ١٥$$

$$n_1 = ١٢$$

$$T = \frac{1}{\frac{1}{10} + \frac{1}{12}} \times \sqrt{\frac{2(6)10 + 2(7,5)12}{2 - 10 + 12}} = 1,85 = \frac{5-}{2,7} =$$

٦- الاستنتاج: بما أن قيمة (ت) المحسوبة (١,٨٥) أكبر من قيمة (ت) النظرية (٢,٠٦) فإنها لا تقع في منطقة الرفض (المنطقة الحرجة)، ولذلك نقبل فرض العدم القائل أن متوسط الإنتاج في المصنعين متساوى، أى أنه لا يوجد فرق جوهري بين المتوسطين وذلك بمستوى معنوية ٠,٠٥ وأن أى فرق بينهما هو فرق يرجع للصدفة المطلقة أو ينتج عن خطأ المعاينة.

ثانياً: تحليل التباين

Analysis of Variance

(اختبار - ف)

التباين هو أحد مقاييس التشتت التي عرفنا من قبل كيفية حسابها من قيم مفردة أو من جداول التوزيعات التكرارية. ولقد تأكدت أهمية التباين في الدراسات والبحوث التي تقوم على أساس إحصائي كمي من حيث أنه المقياس الذي يوضح مدى التجانس أو الاختلاف لثلاث عينات أو أكثر ومدى صلتها بالمجتمع الإحصائي الذي تمثله. ولذا تعد طريقة تحليل التباين أشهر وأهم طرق التحليل الإحصائي للبيانات بعامة.

وفي الفصل السابق والقسم الأول من هذا الفصل تمت معرفة كيفية تحليل واختبار الفروق بين متوسطى عینتين للحكم على خصائص مجتمعيهما، أى أنه تحليل يتعلق بمجتمعين أو عینتين فقط. ولقد اعتمدنا فى تحديد مستويات المعنوية أو الدلالة الاحصائية لاختبار الفروق بين المتوسطات على قيم «ز» فى حالة معرفة تباين المجتمع وذلك لقبول أو رفض الفرض الموضوع، أو على قيم «ت» فى حالة عدم معرفة تباين المجتمع أى على أساس معرفة تباين العينة. والسؤال الآن هو

كيف يتصرف الباحث إذا كان لديه أكثر من عینتین ؟ وللإجابة على هذا التساؤل فان الباحث سيجتهد للقیام بعمليات حسابیة بین متوسطات العینات كل عینتین على حده. وتحدد العمليات الحسابیة على أساس الصیغة التالیة:

$$\frac{n(n-1)}{2}$$

حيث n هي عدد العینات المطلوب حساب الفروق لمتوسطاتها.

وللتخلص من كثرة العمليات الحسابیة يمكن المقارنة بین متوسطات أكثر من عینتین باستخدام طريقة أخرى وهي تحليل التباين للعینات^(١) "Analysis of Variance"

تحليل التباين:

يعتمد تحليل التباين أساساً على حساب التباين بین العینات Variance Samples between والتباين داخل كل العینات مجتمعة Samples Variance. أما المقياس المستخدم للحكم على مستوى معنویة أو دلالة الفروق بین متوسطات العینات فهو ما يطلق عليه بقيمة F . وتقاس قيم F النظرية من جداول خاصة موضوعة لهذا الغرض عن طريق تحديد درجات الحرية لكل تباين على حدة بین العینات وداخل العینات. ودرجات الحرية لتباين بین العینات عبارة عن $h - 1$ حيث « h » هي عدد العینات. أما درجات الحرية لتباين داخل العینات فهي « $y - 1$ » (هـ) حيث « y » هي العدد الكلى للمفردات. فمثلاً إذا كان هناك ٦ عینات وكل عينة مكونة من ١٠ مفردات (قياسات)، فإن درجات الحرية في هذه الحالة هي:

$$د ب درجات الحرية لتباين بین العینات = (h - 1) = (1 - 6) = 5$$

$$د د درجات الحرية لتباين داخل العینات = (y - 1) = (10 \times 6) - 6 = 54$$

$$54 = 6 - 60$$

(١) تسمى طريقة تحليل التباين في بعض الأحيان باختبار نسبة F ratio test نسبة إلى عالم الاحصاء Fisher مكتشف هذه الطريقة.

وهناك عدة شروط أساسية يجب أن تتوافر عند استخدام طريقة تحليل التباين لعدة عينات وبغيرها تكون نتائج هذه الطريقة مضللة:

١- أن يكون توزيع مفردات أو قيم العينات متصفاً بصفة الاعتدالية أو أن انحرافها عن التوزيع المعتدل بسيطاً.

٢- أن يكون التباين لقيم المجموعات متجانساً أو متماثلاً، أى لا توجد فروق بين تباين العينات المقارنة الا نتيجة للصدقة وذلك عن طريق مقارنة تباينات العينات.

٣- أن تكون العينات المطلوب تطبيق تحليل التباين عليها ذات ظروف واحدة أو متجانسة.

وتجدر الإشارة هنا إلى أنه يجب قبل بدئ وضع العمليات الحسابية لتحليل التباين أن يوضع فرض لاختباره بهذا المقياس وهو فى هذه الحالة فرض العدم الذى ينص على أنه لا توجد فروق بين متوسطات وتباين العينات الداخلة فى التحليل. ولاختيار هذا الفرض تحسب قيمة (ف) بين العينات وداخل العينات باتباع الخطوات التالية:

١- حساب التباين بين العينات عن طريق حساب المربعات بين العينات.

٢- حساب التباين داخل العينات عن طريق حساب المربعات داخل العينات.

٣- حساب نسبة (ف) عن طريق قسمة التباين الأكبر على التباين الأصغر.

٤- حساب درجات الحرية لاستخدامها فى الكشف عن مستوى الدلالة أو المعنوية الاحصائية لنسبة (ف) المحسوبة ومايقابلها من نسبة (ف) النظرية. فاذا كانت قيمة (ف) المحسوبة أقل من نظيرتها فى الجدول حسب مستوى المعنوية أو درجة الثقة المطلوبة فان الفرض يرفض بمعنى أن هناك فروق جوهرية بين متوسطات العينات. والجدول التالى يوضح صورة النتائج النهائية لهذا الاختبار.

جدول رقم (٩-١)

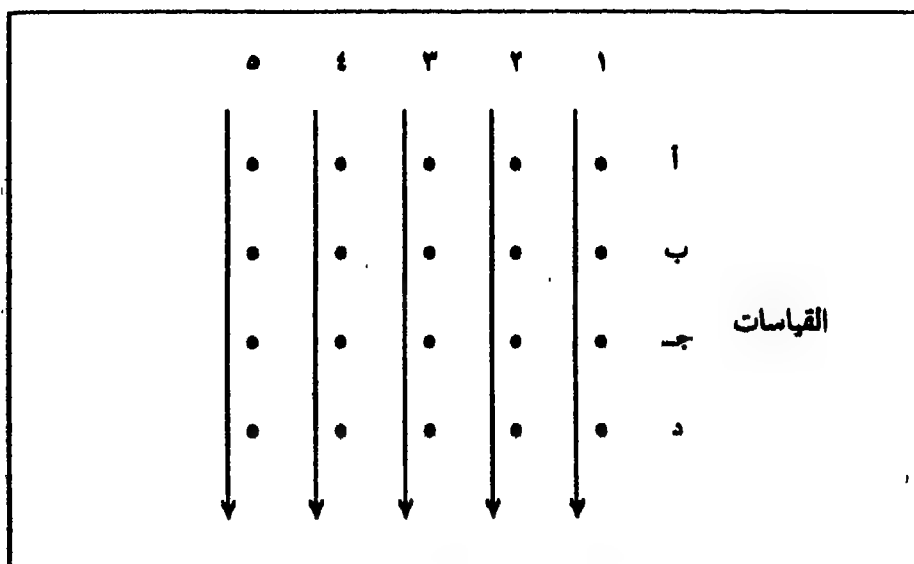
خطوات حساب تحليل التباين لأكثر من عینتين

مصدر التباين	مجموع المربعات	درجة	ف
١- بين العينات	$\frac{\text{مجموع مربع مجموع كل مجموعة}}{ن}$ $- \text{مجموع مربعات القيم}$	عدد العينات ١-	مجموع المربعات مقسوماً على التباين الأكبر مقسوماً على التباين
٢- داخل العينات	الفرق بين ١، ٣	عدد القيم عدد	مجموع المربعات مقسوماً على درجات التباين
٣- المجموع الكلي	مجموع مربعات القيم - $\text{مجموع مربعات القيم}$	مجموع درجات	

حيث أن (ن م) هي عامل التصحيح وهو يساوي أيضاً $\frac{\text{مجموع س}^2}{ن}$ حيث
مجموع س^٢ هو مجموع كل قيم المفردات للعينات و ن هي العدد الكلي لكل
مفردات العينات.

التحليل الاحادي التصنيف للتباين:

في هذا النوع من التحليل يمكن تقسيم الاختلافات الكلية إلى مصدرين
هما: الاختلافات التي ترجع إلى قياسات العينات والأخرى ترجع إلى الأخطاء
التجريبية Experimental Error. وتتكون البيانات هنا من عدد من العينات المستقلة
بكل منها مجموعة من القياسات (شكل رقم ٩-١).



شكل رقم (٩-١)

طريقة جمع البيانات للتحليل الاحادى للتباين

وكما هو واضح من الجدول رقم (٩-١) فانه لحساب قيمة (ف) نتبع الخطوات السابقة، والتي تعرف بطريقة التحليل الأحادى للتباين فى الاجابة على المثال الآتى:

فى مقارنة لمعرفة تباين الإنتاج الزراعى لمحصول الأرز فى خمس مناطق حسب إنتاج كل منها لمدة ٢٠ سنة كما هو موضح فى الجدول رقم (٩-٢). والمطلوب اختبار الفرض القائل بعدم وجود فروق بين متوسطات الإنتاج وذلك بمستوى معنوية ٠,٠٥.

جدول رقم (٩-٢)
الإنتاج السنوى لمحصول الأرز فى المناطق
أ، ب، ح، د، هـ (طن/ فدان) لفترة ٢٠ سنة

١	٢	ب	ب	ج	ج	د	د	هـ	هـ
٢,٥	٦,٢٥	٢,٠	٤,٠٠	٢,١	٤,٤١	٢,٧	٧,٢٩	٣,٤	١١,٥٦
٢,٤	٥,٧٦	٤,٩	٢٤,٠١	١,٥	٢,٢٥	٣,٠	٩,٠٠	٤,٤	١٩,٢٤
٢,٦	٦,٧٦	٣,٣	١٠,٨٩	١,٣	١,٦٩	٣,٢	٠,٢٤	٤,٠	١٦,٠٠
١,٧	٢,٨٩	٣,٨	١٤,٤٤	١,٢	١,٤٤	٢,٤	٥,٧٦	٤,٩	٢٤,٠١
٢,٠	٤,٠٠	٤,٠	١٦,٠٠	١,٦	٢,٥٦	٣,٤	١١,٥٩	٤,٩	٢٤,٠١
٢,٦	٦,٧٦	٣,٥	١٢,٢٥	٢,٢	٤,٨٤	٢,٨	٧,٨٤	٤,٠	١٦,٠٠
٢,٢	٤,٨٤	٢,٦	٦,٧٦	١,٨	٣,٢٤	٣,٧	١٣,٦٩	٤,٧	٢٢,٠٩
٥,١	٢٦,٠٠	٣,٥	١٢,٢٥	١,٧	٢,٨٩	٣,٥	١٢,٢٥	٤,٠	١٦,٠٠
٣,٢	١٠,٢٤	٣,٣	١٠,٨٨	٢,٧	٧,٢٩	٤,٧	٢٢,٠٩	٣,٤	١١,٥٦
٢,٢	٤,٨٤	٣,٨	١٤,٤٤	٣,٩	١٥,٢١	١,٩	٣,٦١	٢,٥	٦,٢٥
٣,٤	١١,٥٦	٣,٤	١١,٥٦	٢,٦	٦,٧٦	٢,٠	٤,٠٠	٧,٢	٥١,٨٠
٢,٠	٤,٠٠	١,٥	٢,٢٥	٢,٥	٦,٢٥	٢,٠	٤,٠٠	٤,٥	٢٠,٢٥
٢,٠	٤,٠٠	٢,٨	٧,٨٤	٢,١	٤,٤١	٢,٠	٤,٠٠	٣,٧	١٣,٦٩
٢,٣	٥,٢٩	٢,٠	٤,٠٠	٢,٢	٤,٨٤	٣,٦	١٢,٩٦	٥,٠	٢٥,٠٠
٣,٠	٩,٠٠	٢,٤	٥,٧٦	٢,٣	٥,٢٩	٢,٥	٦,٢٥	٦,٤	٤١,٠٠
٢,٥	٦,٢٥	٢,٧	٧,٢٩	١,٤	١,٩٦	١,٧	٢,٨٩	٤,٩	٢٤,٠١
١,٨	٣,٢٤	٢,٣	٥,٢٩	١,٣	١,٦٩	٢,٨	٧,٤٨	٣,٥	١٢,٢٥
٢,٤	٥,٧٦	٢,٩	٨,٤١	١,٤	١,٩٦	٢,٦	٦,٧٦	٩,٧	٩٤,٠٠
٢,٩	٨,٤١	٢,٠	٤,٠٠	١,١	١,٢١	٢,٦	٧,٧٦	٥,٠	٢٥,٠٠
٢,٢	٤,٨٤	٤,٠	١٦,٠٠	١,٥	٢,٢٥	٣,٠	٩,٠٠	٣,٢	١٠,٢٤
٥١,٠٠	٤٠,٦٥	٦٠,٧	١٩٨,٣٣	٣٨,٤	٨٢,٤٤	٥٦,١	١٦٧,٧٩	٩٣,٣	٤٨٤,١٧

$$\text{المجموع (مجم س)} = 93,3 + 56,1 + 38,4 + 60,7 + 51,0 = 299,5$$

$$\text{عامل التصحيح} = \frac{\sum (\text{مجم س})}{n} = \frac{\sum (299,5)}{100} = 897,04$$

$$\text{المجموع الكلي للمربعات} = 167,79 + 82,44 + 198,33 + 140,69 = 589,25$$

$$176,19 = 897,04 - 1073,23 = 897,04 - 484,17 +$$

$$\text{مجموع المربعات بين العينات} = \frac{1}{2} [\sum (38,4) + \sum (60,7) + \sum (51,0)]$$

$$82,9 = 897,04 - [\sum (93,3) + \sum (56,1) +$$

$$\text{مجموع المربعات داخل العينات} =$$

$$\text{المجموع الكلي للمربعات} - \text{مجموع المربعات بين العينات}$$

$$93,29 = 82,90 - 176,19 =$$

$$\text{درجات الحرية بين العينات} = \text{عدد العينات} - 1$$

$$4 = 5 - 1 =$$

$$\text{درجات الحرية داخل عينات} = \text{عدد المفردات} - \text{عدد العينات}$$

$$95 = 100 - 5 =$$

$$\text{التباين بين العينات} = \frac{\text{مجموع المربعات بين العينات}}{\text{درجات الحرية بين العينات}}$$

$$20,73 = \frac{82,9}{4} =$$

$$\text{التباين داخل العينات} = \frac{\text{مجموع المربعات داخل العينات}}{\text{درجات الحرية داخل العينات}}$$

$$0,982 = \frac{93,29}{95} =$$

$$\text{نسبة ف} = \frac{\text{التباين الأكبر}}{\text{التباين الأصغر}} = \frac{\text{التباين بين العينات}}{\text{التباين داخل العينات}}$$

$$21.11 = \frac{20.73}{0.982} =$$

وفى العادة توضح النتائج السابقة فى جدول كمايلى:

مصدر التباين	مجموع المربعات	درجات الحرية	التباين	ف
بين العينات	٨٢,٩٠	٤	٢٠,٧٣	٢١,١١
داخل العينات	٩٢,٢٩	٩٥	٠,٩٨٢	
المجموع	١٧٦,١٩	٩٩		

ويستخدم جدول «ف» النظرية (F- Tables) وهو عبارة عن جدول لحساب نسبة التباين بدرجات بين العينات وداخل العينات وبمستوى معوية ٠,٠٥ وفى هذا الجدول تكون درجات الحرية الأفقية خاصة بدرجات الحرية للتباين الأصغر ودرجات الحرية الرأسية خاصة بدرجات الحرية للتباين الأكبر. وفى المثال بين أيدينا نجد أن نسبة «ف» لدرجات حرية (٤) بين العينات (ذات التباين الأكبر)، (٩٥) داخل العينات ذات التباين الأصغر عند مستوى دلالة ٠,٠٥ هى ٢,٤٩ تقريباً. وبما أن قيمة ف المحسوبة أكبر من القيمة النظرية فانه يمكن رفض فرض العدم الأساسى القائل بعدم وجود فروق جوهرية بين متوسطات إنتاج الأرز فى المناطق الخمسة، أو بمعنى آخر نقبل الفرض البديل وهو أنه توجد فروق جوهرية ذات دلالة إحصائية بين متوسطات الإنتاج. أى أن هناك احتمال مقداره ٠,٩٥ أن لا تكون الفروق بين المتوسطات قد حدثت بفعل الصدقة أو بطريقة عشوائية.

مثال آخر:

فى تجربة لدراسة المحتوى المائى (نسبة الرطوبة) لأربعة أنواع من التربة فى أحد المناطق أخذ من كل نوع عدد من العينات وتم استخلاص النتائج الموضحة فى

الجدول (٣-٩)، فهل يمكن الحكم بأن متوسطات المحتوى المائي (نسبة الرطوبة) للأربعة أنواع من التربة متساوية وذلك بمستوى دلالة ٠,٠١، ٠,٠٥، ٠,٠١؟

جدول رقم (٣-٩)

المحتوى المائي (نسبة الرطوبة) لأربعة أنواع من التربة في منطقة ما

أ	ب	ب ^٢	ج	ج ^٢	د	د ^٢
٢٣	٢٥	٦٢٥	١٧	٢٨٩	١٥	٢٢٥
٢٦	٢٨	٧٨٤	١٩	٣٦١	١٤	١٩٦
٣٢	٢٥	٦٢٥	١٢	١٤٤	١٢	١٤٤
٣٣	٢٤	٥٧٦	١٦	٢٥٦	١٣	١٦٩
	١٨	٣٢٤			١١	١٢١
١٢٤	١٢٠	٢٩٣٤	٦٤	١٠٥٠	٦٥	٨٥٥
٤	٥		٤		٥	
٣١	٢٤		١٦		١٣	
المجموع						
عدد المفردات						
(ن)						
المعوسط						

$$١- \text{المجموع (س ن)} = ٩٤ + ٩٠ + ٦٤ + ٦٥ = ٣٧٣$$

$$٢- \text{عامل التصحيح} = \frac{\text{مجم (س ن)}}{ن} = \frac{\sum (٣٧٣)}{١٨} = ٧٧٢٩,٣٩$$

$$٣- \text{مجموع المربعات بين العينات} =$$

$$٧٧٢٩,٣٩ - \left[\frac{\sum (٦٥)}{٥} + \frac{\sum (٦٤)}{٤} + \frac{\sum (١٢٠)}{٥} + \frac{\sum (١٢٤)}{٤} \right]$$

$$٩٨٧,٦١ = ٧٧٢٩,٢٩ - ٨٥٩٣,٠٠ =$$

$$- ٤ - \text{مجموع المربعات الكلية} = [٨٥٥ + ١٠٥٠ + ٢٩٣٤ + ٣٨٧٨] - \\ ٩٨٧,٦١ = ٧٧٢٩,٣٩$$

$$- ٥ - \text{مجموع المربعات داخل العينات} = \text{مجموع المربعات الكلية} - \text{مجموع المربعات بين العينات}$$

$$١٢٤,٠٠ = ٨٦٣,٦١ - ٩٨٧,٦١ =$$

$$- ٦ - \text{درجات الحرية للمكونات السابقة:}$$

$$\text{درجات الحرية بين العينات} = \text{عدد العينات} - ١ = ٤ - ١ = ٣$$

$$\text{درجات الحرية الكلية} = \text{عدد المفردات} - ١ = ١٨ - ١ = ١٧$$

$$\text{درجات الحرية داخل العينات} = \text{عدد المفردات} - \text{عدد العينات}$$

$$١٤ = ٤ - ١٨ =$$

$$(\text{وهي تساوى أيضاً درجات الحرية الكلية} - \text{درجات الحرية بين العينات}) \\ (١٤ = ٣ - ١٧ =$$

$$- ٧ - \text{التباين بين العينات} = \frac{٨٦٣,٦١}{٣} = ٢٨٧,٩$$

$$\text{التباين داخل العينات} = \frac{١٢٤,٠٠}{١٤} = ٨,٩$$

$$- ٨ - \text{قيمة (ف)} = \frac{\text{التباين الأكبر}}{\text{التباين الأصغر}} = \frac{٢٨٧,٩}{٨,٩} = ٣٢,٥$$

وبتلخيص البيانات في صورة جدولية نحصل على جدول تحليل التباين التالي:

جدول رقم (٩-٤)
تحليل التباين لمتوسطات المحتوى المائي (نسبة الرطوبة)
لأربعة أنواع من التربة

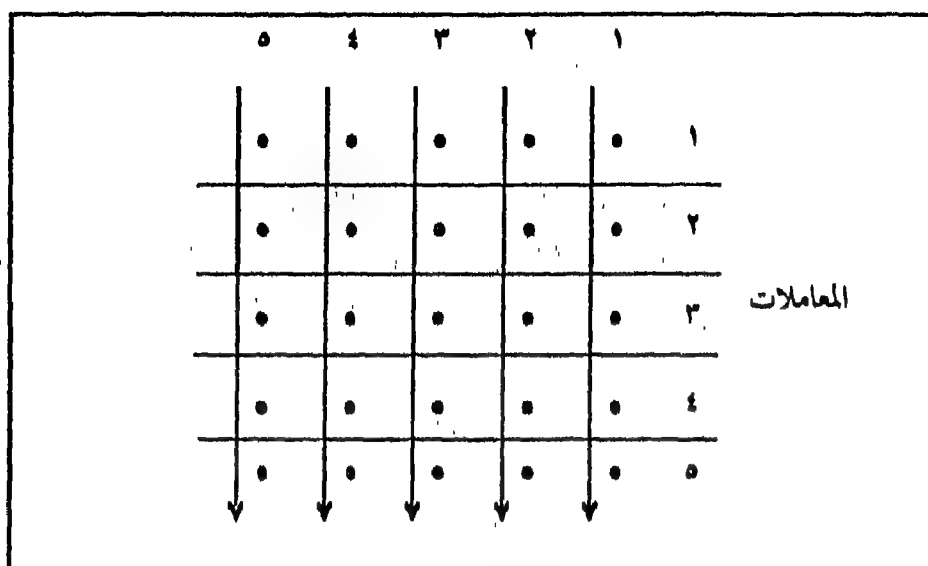
مصدر التباين	مجموع المربعات	درجات الحرية	التباين	قيمة (ف)
بين العينات	٨٦٣,٦١	٣ = ١ - ٤	٢٨٧,٩	٣٢,٥
داخل العينات	١٢٤,٠٠	١٤ = ٤ - ١٨	٨,٩	
المجموع	٩٨٧,٦١	١٧ = ١ - ١٨		

ولاختبار الفرض القائل بتساوى المتوسطات الأربعة نقارن قيمة (ف) المحسوبة بقيمة (ف) النظرية عند درجات حرية ٣، ١٤ ومنها يظهر من جداول ف - أن قيمة (ف) لمستوى دلالة ٠,٠٥ = ٢,٣٤ وقيمة (ف) لمستوى دلالة ٠,٠١ = ٥,٥٦. وحيث أن قيمة ف المحسوبة (٣٢,٥) أكبر من قيمة (ف) في الجدول فإننا نرفض فرض العدم أى أن الاختلافات بين متوسطات العينات (أنواع التربة) هي اختلافات جوهرية حقيقية لا ترجع إلى الصدقة (أى أنها أكبر من الاختلافات العشوائية)، وذلك باحتمال ٠,٠٥ أو ٠,٠١ وبمعنى آخر أن متوسطات المحتوى المائي (نسبة الرطوبة) للأنواع الأربعة من التربة غير متساوية.

التحليل الثنائي التصنيف للتباين Two - way Analysis of Variance

إنفرض من المثال السابق كيفية تحليل التباين لأكثر من مجموعتين بطريقة التحليل الأحادى التصنيف أو التحليل بالتصنيف فى اتجاه واحد. وهناك طريقة لتحليل التباين تعرف باسم التحليل الثنائي للتباين. وفى هذا النوع من التحليل يمكن تقسيم الاختلافات الكلية إلى ثلاث مصادر: اختلافات ترجع إلى العينات، اختلافات ترجع إلى المعاملات Treatments، واختلافات ترجع إلى الأخطاء

التجريبية Experimental. وتتكون البيانات هنا من عدد من العينات المستقلة بكل منها عدد من المفردات (المعاملات) كما في الشكل رقم (٩-٢). ويوضح المثال التالي كيفية تطبيق طريقة الشاى للتباين وخطوات العمل اللازمة معها:



شكل رقم (٩-٢)

طريقة أخذ البيانات لاختيارها بأسلوب التحليل الشاى للتباين

مثال:

نفترض أننا سحبنا عينة عشوائية طبقية مكونة من ثلاث مجموعات من الأحواض الزراعية فى زمام أحد المراكز الإدارية بهدف مقارنة الإنتاجية الزراعية بين أنواع (طبقات) التربة التى تمثلها هذه المجموعات وهى التربة الطينية السوداء، والتربة الطينية الصفراء، والتربة الرملية. فإذا كانت كل مجموعة مكونة من عشرة أحواض تم إختيارها عشوائياً ثم رصدت لها كميات الإنتاج لمحاصيل الحبوب (بالطن) كما فى الجدول رقم (٩-٣)، فهل نقبل الفرض القائل بعدم وجود فروق بين متوسطات الإنتاج لهذه الأنواع الثلاثة من التربة.

جدول رقم (١٢-٣)
الإنتاجية الزراعية لحاصل الحبوب (بالطن)
لمجموعة من الأحواض الزراعية في ثلاثة أنواع من التربة

المجموع		التربة الرملية		التربة الطينية		التربة الطينية		
٢ (س ن)	٢ (س ن)	ج ٢	ج	ب ٢	ب	أ ٢	أ	
٢٤	٢٤	٢٤	٢٤	٢٤	٢٤	٢٤	٢٤	
٢٧	٢٧	٢٧	٢٧	٢٧	٢٧	٢٧	٢٧	
٢١	٢١	٢١	٢١	٢١	٢١	٢١	٢١	
٢٢	٢٢	٢٢	٢٢	٢٢	٢٢	٢٢	٢٢	
٢٦	٢٦	٢٦	٢٦	٢٦	٢٦	٢٦	٢٦	
١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	
٢٥	٢٥	٢٥	٢٥	٢٥	٢٥	٢٥	٢٥	
٢٩	٢٩	٢٩	٢٩	٢٩	٢٩	٢٩	٢٩	
٢٦	٢٦	٢٦	٢٦	٢٦	٢٦	٢٦	٢٦	
٢٤	٢٤	٢٤	٢٤	٢٤	٢٤	٢٤	٢٤	
٢٤٣	٢٤٣	٢٤٣	٢٤٣	٢٤٣	٢٤٣	٢٤٣	٢٤٣	المجموع

$$١- \text{المجموع (م.س.ن)} = ٢٤٣ + ٢٢٢ + ٢١٠ = ٦٧٥$$

$$٢- \text{عامل التصحيح} = \frac{٢(٦٧٥)}{٣٠} = \frac{٢(م.س.ن)}{ن}$$

$$٣- \text{مجموع مربعات بين العينات (المجموعات)} = ١٥١٨٧,٥ - \frac{٤٥٨٥٣}{٣}$$

$$= ٩٦,٨٣$$

٤- مجموع المربعات بين المعاملات

$$= \frac{1}{1} [(210)^2 + (222)^2 + (243)^2] - 15187,5$$

$$= 55,8 = 15187,5 - 15243,3$$

٥- المجموع الكلى للمربعات = ١٥٤٨٩,٠ = ١٥١٨٧,٥ - ٣٠١,٥

٦- نحسب الخطأ التجريبي Experimental Error لمجموع المربعات (الفرق أو الفائض Residual) وهو عبارة عن الفرق بين المجموع الكلى للمربعات ومجموع المربعات بين العينات ومجموع المربعات بين المعاملات كمايلي:

$$\text{خطأ مجموع المربعات} = 301,5 - 96,83 - 55,8 = 148,87$$

٧- نحسب درجات الحرية للتحليل كمايلي:

درجات الحرية بين المعاملات = عدد المعاملات (بين الصفوف) - ١

$$= 10 - 1 = 9$$

درجات الحرية بين العينات (بين الأعمدة) = عدد العينات - ١

$$= 3 - 1 = 2$$

درجات الحرية للمفردات الكلية (ن) = عدد المفردات (ن) - ١

$$= 30 - 1 = 29$$

درجات الحرية للخطأ (الفرق) = درجات الحرية للمفردات - درجات حرية

التباين بين العينات - درجات الحرية بين المعاملات (بين الصفوف)

$$= 29 - 2 - 9 = 18$$

أو

= (درجة الحرية بين العينات) × (درجات الحرية بين المعاملات)

$$= 2 \times 9 = 18$$

ويمكن تلخيص النتائج السابقة فى الجدول التالى:

جدول رقم (٩-٤)

جدول التحليل الثنائي للتيابن 2 WAY - ANOVA

قيمة ف	التيابن	درجات الحرية	مجموع المربعات	مصدر التباين
٥,٨٥	٤٨,٤١٥	٢	٩٦,٨٣	بين العينات
		٩	٥٥,٨٠	بين المعاملات
		١٨	١٤٨,٨٧	اخطأ (داخل المعاملات)
	٨,٢٧	٢٩	٣٠١,٥٠	المجموع الكلي

وبالبحث في جداول (ف) لإيجاد قيمتها النظرية عند درجات الحرية ٢ بين المجموعات و ١٨ للخطأ عند مستوى دلالة ٠,٠٥ نجد أن $F_{0.05} = 3.55$. وحيث أن قيمة (ف) المحسوبة في المثال (٥,٨٥) أكبر من القيمة النظرية عند مستوى الدلالة ٠,٠٥ فإننا يجب أن نرفض الفرض القائل أنه لا توجد فروق جوهرية بين متوسطات المجموعات الثلاث من التربة من حيث الإنتاجية الزراعية للحبوب. وهذا يعنى أن هناك احتمالاً قدرة ٠,٩٥ أن الفروق بين الإنتاجية الزراعية للحبوب لهذه الأنواع من التربة لم تحدث بفعل الصدفة المطلقة أو أنها نتيجة خطأ عشوائى فى المعاينة ولكنها فروق حقيقية لها دلالة إحصائية.

الفصل العاشر
أساليب المقارنة الالاباراميتريية
(اللامعلمية)

الفصل العاشر

أساليب المقارنة اللاباراميتريّة (اللامعلمية)

عرضنا في الفصل السابق الأساليب الكمية الباراميتريّة التي تحلل المعنوية الاحصائية للاختلافات أو الفروق بين بيانات التوزيعات العينية والتوزيع المعتدل للمجتمع، وكان الاهتمام منصّباً على مقارنة قيم المتوسط الحسابي والتباين المحسوبة مع عينتين أو أكثر لمتغير واحد. وذكرنا أيضاً أنه في بعض الحالات لا تكون خصائص أو معالم المجتمع معلومة، وبالتالي لا بد أن تستخدم أساليباً أخرى لإجراء اختبارات المعنوية وتحليل البيانات الأصلية التي لا تتصف باعتدالية توزيعاتها التكرارية حتى بعد تطبيق إحدى الطرق السابق ذكرها لتحويلها إلى توزيعات معتدلة.

وستتناول في هذا الفصل دراسة الأساليب الكمية اللاباراميتريّة (اللامعلمية) التي تستخدم في مجال المقارنة بين مجموعتين أو أكثر من البيانات لمتغير واحد، وهي: مربع كاي، اختبار كولموجوروف - سميرونوف (اختبار «د» اختبار مان - هويتني (اختبار «ي»)، اختبار ويلكوكسون، اختبار كروسكال - واليس (اختبار «هـ»). وتجدر الإشارة إلى أن هذه الاختبارات لا تشترط أن تكون البيانات من بيانات الفترة فقط ولكن يمكن تطبيقها أيضاً على البيانات الاسمية (التصنيفية أو النوعية) والترتيبية سواء كانت في صورة قياسات فردية أو ثنائية ينتج عنها قيم عددية أو تكرارات أو رتب.

أولاً: اختبار مربع كاي (x2)

Chi - Square Test

يستخدم مربع كاي أساساً في قياس مدى التطابق بين توزيعين أحدهما توزيع فعلي لمتغير تم قياسه والآخر توزيع نظري أو متوقع. وعلى ذلك فوجه المقارنة يكون بين مجموعتين من البيانات التكرارية أحدهما فعلية والآخرى نظرية. ويكون الفرض الموضوع هو المتعلق بالفروق أو الاختلافات بين التوزيعات الفعلية أو المشاهدة والتوزيعات المتوقعة للوقوف على معرفة نوع هذه الفروق: هل هي فروق معنوية أو جوهرية، أم أنها مجرد فروق ظاهرية؟ فإذا كانت الفروق حقيقية فإن ذلك يعنى أنها نتيجة لعوامل مسؤولة عنها وليست مرتبطة بعوامل أخرى مسببة لها، أما إذا كانت الفروق غير جوهرية فإن ذلك يعنى أنها نتيجة للصدفة المطلقة أو أنها نتيجة لمجموع العوامل غير المتحكم فيها أو ما يطلق عليه بالاختلافات والأخطاء العشوائية Random Errors. وبصفة عامة فإنه يمكن القول أن تحليل البيانات بواسطة مربع كاي يتم لتحقيق هدفين رئيسيين هما:

١- تحديد دلالة العلاقة بين مجموعتين أو أكثر من التصنيفات بالنسبة إلى خصائص العينة.

٢- تحديد دلالة انحرافات التكرارات الفعلية (المشاهدة) عن التكرارات المتوقعة، أو بمعنى آخر الحكم على مدى ملائمة النموذج المتوقع لتوزيع التكرارات الفعلية عن طريق اتباع اختبار جودة التوفيق Goodness of fit Test.

وكما سبق أن شاهدنا أنه في عديد من المرات لا تتفق النتائج التي نحصل عليها من العينات في جميع الحالات مع النتائج المتوقعة طبقاً لقواعد الاحتمالات. فلو افترض أنه في عينة معينة لوحظ أن مجموعة من الحوادث الممكنة: هـ_١، هـ_٢، هـ_٣، هـ_٤ ... هن تحدث بتكرارات ش_١، ش_٢، ش_٣، ... ش_ن التي تسمى بالتكرارات المشاهدة، وأنه طبقاً لقواعد الاحتمالات فإنه يتوقع أن تحدث بتكرارات ق_١، ق_٢، ق_٣ ... ق_ن والتي تسمى بالتكرارات المتوقعة كما في الشكل التالي:

الحادث	١-هـ	٢-هـ	٣-هـ هـ-ن
التكرار المشاهد	١ش	٢ش	٣ش ش-ن
التكرار المتوقع	١ق	٢ق	٣ق ق-ن

ويتطلب إجراء اختبار «مربع كاي» توفر الشروط الأساسية الآتية:

أولاً: أن لا يقل العدد الكلي للقيم (حجم العينة) عن ٢٠.

ثانياً: أن تكون المشاهدات مستقلة عن بعضها البعض (أى لا يؤثر اختيار أحد المفردات على اختبار المفردات الأخرى).

ثالثاً: أن تكون البيانات المشاهدة فى شكل قيم عددية أو تكرارات قائمة على العد فى كل فئة من الفئات، ولا تكون هذه البيانات فى صورة نسب مئوية على الإطلاق.

رابعاً: إذا كانت العينة تنقسم إلى فئتين فقط فيجب أن لا يقل عدد التكرارات المتوقعة لهما عن ٥ تكرارات. إذا زاد عدد الفئات عن فئتين فيجب أن لا يقل ١/٥ التكرارات المتوقعة، على الأكثر، عن ٥ تكرارات، ولكن يجب أن لا يقل أبداً عدد التكرارات المتوقعة لأية فئة عن تكرار واحد. فإذا لم يتحقق ذلك فى العينات المقارنة فإنه يمكن ادماج فئتين أو أكثر فى فئة واحدة حتى يمكن أن تتم إجراءات المقارنة بأداة مربع كاي.

ولحساب قيمة مربع كاي تستخدم الصيغة الاحصائية الآتية:

$$\text{مربع كاي } (x^2) = \frac{(ش_١ + ق_١)^2}{ق_١} + \dots + \frac{(ش_٣ + ق_٣)^2}{ق_٣}$$

حيث (ش) هي التكرارات المشاهدة (ق) هي التكرارات المتوقعة أو النظرية. وتتراوح قيمة مربع كاي من صفر إلى α ويتوقف توزيعها على درجات الحرية. فإذا كانت النتيجة النهائية (القيمة المحسوبة) لاختبار مربع كاي أقل من نظيرتها في توزيع مربع كاي^(١) في الجداول الاحصائية الخاصة به في مستوى دلالة معين (α) فإننا نقبل فرض العدم أو بمعنى آخر أنه لا يوجد اختلافات أو فروق معنوية بين التوزيعين المشاهد (الفعلي) والمتوقع، أما إذا كانت قيمة مربع كاي المحسوبة أكبر من مثلتها النظرية في جدول توزيع مربع كاي فإننا نرفض فرض العدم، وهذا يعني وجود فروق معنوية أو حقيقية بين التوزيعين أي أن هناك عوامل غير عامل الصدفة لها تأثير على هذه الفروق. وتقوم المقارنة بين قيمة مربع كاي المحسوبة والنظرية، على أساس درجات الحرية ($n - 1$) حيث n هي عدد المجموعات أو الفئات. هذا في حالة التصنيفات الأحادية أما في حالة التصنيفات الثنائية كجداول الاقتران Contingency Tables فتحسب درجات الحرية كما يلي:

$$(d-1) (1-1) \text{ حيث } d = \text{عدد الأعمدة، } v = \text{عدد الصفوف}$$

التحليل الأحادي The one - Sample Case

يتصل هذا التصنيف بالبيانات المشاهدة التي يتم تقسيمها حسب صفة واحدة من الصفات.
مثال:

جمعت عينة عشوائية لعدد ٦٠٠ قطعة صخرية من منطقة شاطئية، ووجد أن ١٨٠ قطعة منها من نوع الحجر الجيري، ١٨٦ من الجرانيت والباقي ٢٣٤ من نوع الحجر الرملي. وذلك على الرغم من أن الأنواع الصخرية الثلاثة متمثلة في المنطقة بمساحات متساوية، والفرض المراد اختباره في هذه الحالة: حيث أن الأنواع

(١) توزيع مربع كاي النظرى غير منتظم، متوسطه يساوى درجات الحرية وتباينه يساوى ضعف درجات الحرية، ويقرب من التوزيع المعتدل كلما زادت درجات الحرية.

الرئيسية للصخور مساحتها متساوية في المنطقة فإننا نتوقع أن يكون لكل نوع منها عدداً من القطع مماثلاً للنوع الآخر. والفرض البديل اختباره في هذه الحالة: حيث أن الأنواع الرئيسية للصخور مساحتها متساوية في المنطقة فإننا نتوقع أن يكون لكل نوع منها عدداً من القطع مماثلاً للنوع الآخر. والفرض البديل لذلك هو وجود فروق جوهرية في عدد القطع الصخرية لكل نوع من أنواع الصخور. وحيث أن أعداد القطع قد جمع في عينة واحدة فإنه ليس من الصواب أن نتوقع تساوى أعداد القطع لكل نوع، ولكن بما أن العينة عشوائية فإن أعداد القطع لا تبعد كثيراً عن التساوى من بعضها البعض، وفي هذه الحالة فإن مربع كاي يستخدم لتقدير الاحتمال القائل أن العينة قد أخذت من مجتمع تتساوى فيه أعداد القطع لكل نوع صخرى مع العلم بأن مستوى الدلالة المطلوب هو ٠,٠١ ويمكن ترتيب البيانات السابقة كما هي الحال في جدول المقارنة التالي:

جدول رقم (١٠ - ١)

حساب مربع كاي بطريقة التحليل الأحادي

الحادث	عدد القطع المجموعة (التكرارات المشاهدة) ش	عدد القطع المتوقعة (ق)
حجر جيري	١٨٠	٢٠٠
جرانيت	١٨٦	٢٠٠
حجر رملي	٢٣٤	٢٠٠

$$\text{مربع كاي} = \frac{\sum \frac{(ش_i - ق_i)^2}{ق_i}}{ق_i} = \frac{(ش_1 - ق_1)^2}{ق_1} + \frac{(ش_2 - ق_2)^2}{ق_2} + \frac{(ش_3 - ق_3)^2}{ق_3}$$

$$= \frac{(١٨٠ - ٢٠٠)^2}{٢٠٠} + \frac{(١٨٦ - ٢٠٠)^2}{٢٠٠} + \frac{(٢٣٤ - ٢٠٠)^2}{٢٠٠}$$

$$5,78 + ,98 + 2 =$$

$$8,76 =$$

$$\text{درجات الحرية} = \text{عدد الفئات} - 1 = 3 - 1 = 2$$

وبما أن قيمة مربع كاي من التوزيع النظري المدون في الجداول الخاصة به تساوي 9,71 لدرجة حرية 2 ومستوى دلالة 0,1، فمعنى ذلك أنه لا بد من قبول فرض العدم القائل: أنه يتوقع جمع عدد متساوي للقطع الصخرية من كل نوع من أنواع الصخور في المنطقة بمستوى دلالة 0,1، وبناء عليه فإن الفرق بين أعداد القطع في العينة هو فرق يرجع إلى الصدفة المحضة أو نتيجة خطأ المعاينة.

التحليل (التصنيف) الثنائي The Two Sample Case

لاحظنا في التحليل الأحادي أن الاختبار للفرض الموضوع كان يتعلق بالفروق بين التكرارات المشاهدة والتكرارات المتوقعة. أما في حالة التحليل الثنائي بمربع كاي فإن الفرض يتعلق باختبار الاستقلال بين صفتين أو تصنيفين أو اختبار عدم وجود علاقة بين هاتين الصفتين أو التصنيفين. وبحسب مربع كاي في حالة وجود عينتين وفئتين وعلى النحو التالي:

$$\frac{n \left[\frac{a}{n} - \left(\frac{a+b}{n} \right) \left(\frac{a+c}{n} \right) \right]^2}{(a+b)(c+a)(a+c)(b+c)}$$

حيث n = العدد الكلي لمفردات العينتين (المجموع الكلي للتكرارات)، a ، b ، c ، d = التكرارات لكل فئة من فئتي العينتين داخل «جدول الاقتران» كما هي الحال في الجدول التالي:

(ب + أ)	ب	أ
	د	ج
(ج + د)	(ب + د)	(أ + د)

وتحسب درجات الحرية لهذا الاختبار من عدد الأعمدة وعدد الصفوف في الجدول كما يلي :

$$(د - ١) \times (ص - ١) \text{ حيث } د = \text{عدد الأعمدة، ص} = \text{عدد الصفوف}$$

مثال (٥) :

أخذت عينة لأسبوع واحد من سجلات إدارة مرور لمن تقدموا لامتحان قيادة السيارات فوجد أنه من بين ١٠٠ من النساء اجتاز منهن الامتحان لأول مرة ٧٥ ورسب ٢٥. وأن من بين ١٥٠ رجلاً اجتاز منهم الامتحان لأول مرة ٦٠ ورسب ٩٠، فإذا أخذت العينة السابقة على أساس أنها تمثل قطاعاً عرضياً لمجتمع المتقدمين لامتحان قيادة السيارات، فإن عدد النساء والرجال، كل على حدة، يمكن أن يعامل على أنه عينة عشوائية يهدف تحليلها بمربع كاي إلى معرفة ما إذا كانت هناك اختلافات جوهرية بين مقدرة كل من النساء والرجال على القيادة واجتيازهم اختبار القيادة في أول محاولة بمستوى ثقة ٩٥٪. وجدول الاقتران التالي يوضح بيانات هذا المثال.

جدول رقم (١٠ - ٢)

أعداد المتقدمين لامتحان قيادات السيارات

حسب النوع ونتيجة الامتحان

النوع	ناجح	راسب	المجموع
نساء	٧٥ (أ)	٢٥ (ج)	١٠٠
رجال	٦٠ (ب)	٩٠ (د)	١٥٠
المجموع	١٣٥	١١٥	٢٥٠

ومن جدول الاقتران السابق يمكن إيجاد قيمة مربع كاي كما يلي :

$$\frac{n \left[\frac{n}{2} - (d - b - c) \right]}{(a + b)(c + d)(a + c)(b + d)} = \text{مربع كاي}$$

$$\frac{2 \left[\frac{200}{2} - (25 \times 60 - 90 \times 75) \right] 200}{(90 + 60)(25 + 75)(90 + 25)(60 + 75)} =$$

$$\frac{2 \left[125 - (1500 - 6750) \right] 200}{150 \times 100 \times 115 \times 135} =$$

$$28,196 = \frac{2 \left[5125 \right] 200}{232875000} =$$

$$\text{وحيث أن درجات الحرية} = (d - 1) \times (c - 1)$$

$$1 = (1 - 2) \times (1 - 2) =$$

فإن قيمة مربع كاي من توزيعها في الجدول بدرجة حرية ١ ومستوى ثقة ٩٥ ، (مستوى دلالة ٠,٠٥) هي ٣,٨٤. وبما أن قيمة مربع كاي المحسوبة هي ٢٨,١٩٦ أكبر من القيمة النظرية لمربع كاي فإننا نرفض فرض العدم. ومعنى ذلك أن الاختلاف في مقدرة كل من النساء والرجال في اجتياز امتحان القيادة في أول محاولة لا يرجع إطلاقاً إلى الصدفة في العينة، بل أن بيانات العينة تعكس بكل دقة الاختلاف الحقيقي والجوهرى بين مقدرة النساء والرجال في مجتمع السائقين.

ومما تجدر الإشارة إليه هنا أن معادلة «مربع كاي» المستخدمة لتحليل عينة واحدة يمكن تطبيقها أيضاً في تحليل بيانات عينتين أو أكثر يمكن تقسيمها إلى أكثر من فئتين.

مثال (٢):

أجريت دراسة على عينة عشوائية من ١٥٠ سفاهاً «منحدرأ» طبيعياً أخذت من

ثلاث مناطق متجاورة تتصف كل منها بتركيب صخري متميز (مارل فى المنطقة الأولى، حجر جبرى فى المنطقة الثانية، حجر رملى فى المنطقة الثالثة) لتحديد ما إذا كان هناك اختلاف جوهري بين درجات انحدار السفوح ونوع التركيب الصخري الذى تنشأ فوقه هذه السفوح. وقد وضع فرض العدم فى هذه الحالة وهو أنه لا يوجد اختلاف جوهري بين درجات انحدار السفوح الطبيعية التى تنشأ فوق أنواع مختلفة من التركيب الصخري، وذلك فى مقابل الفرض البديل القائل بأنه على افتراض أن التاريخ الجيولوجى والغطاء النباتى ومقدار الاتجاه للسفوح متشابهة فى المناطق الثلاث فإنه يمكن أن نستنتج أن الاختلافات الجوهريّة فى درجات الانحدار ما هى إلا نتيجة واختلاف طبيعة التركيب الصخري نفسه، وأن أى اختلاف آخر إنما يرجع إلى الصدفة وذلك فى مستوى دلالة ٠,٠٥ ولاختبار الفرض السابق فإن أعداد السفوح لكل نوع من أنواع التركيب الصخري (أى فى المناطق الثلاث، فى العينة قد قسمت إلى خمس فئات حسب درجة الانحدار، ووضعت البيانات الخاصة بذلك فى جدول الإقتران التالى:

جدول رقم (١٠ - ٣): التوزيع التكرارى لأعداد السفوح الطبيعية فوق ثلاثة أنواع مختلفة من التركيب الصخري

الجموع	التكرارات (عدد السفوح)			الفئات
	حجر رملى	حجر جبرى	مارل	(درجة الانحدار)
٣٥	٥	٩	٢١	صفر -
٣٦	٧	٢١	١٨	- ٥
٣٣	٩	١٥	٩	- ١٠
٣٢	١١	١٣	٨	- ١٥
١٤	٨	٥	١	٢٠ إلى أقل من ٢٥
١٥٠	٤٠	٥٣	٥٧	الجموع

فإذا كانت توقعاتنا بأن نسبة عدد السفوح في كل فئة من فئات درجة الانحدار يتناسب مع العدد الكلي السفوح فمعنى ذلك أنه من المفروض أن من مجموع ٥٧ سفحاً طبيعياً تكونت فوق صخور مارلية ٢٣,٢ ٪ منها (أى ١٣,٣ سفحاً) ذات انحدارات تتراوح درجتها بين الصفر وأقل من ٥ درجات، ٢٤ ٪ منها (أى ١٣,٧ سفحاً) ذات انحدارات تتراوح درجتها بين أكثر من ٥ درجات وأقل من ١٠ درجات ٢٢ ٪ منها (أى ١٢,٢ سفحاً) ذات انحدارات تتراوح درجتها بين أكثر من ١٥ درجة وأقل من ٢٠ درجة، وأخيراً ٩,٣ ٪ منها (أى ٥,٣ سفحاً) ذات انحدارات تتراوح درجتها بين أكثر من ٢٠ درجة وأقل من ٢٥ درجة. وبالمثل يمكن إيجاد العدد المتوقع للتكرارات داخل كل فئة من الفئات للأنواع الثلاثة من التركيب الصخرى كما فى الجدول التالى :

جدول رقم (١٠ - ٤) : حساب التكرارات المتوقعة من البيانات

المشاهدة لعدد ١٥٠ سفحاً طبيعياً

الانحدار	التكرارات (عدد السفوح)			الفئات (درجة الانحدار)
	صخر رملى	صخر جبرى	مارل	
٢٥	$\frac{35 \times 40}{150}$ (٩,٣)	$\frac{35 \times 53}{150}$ (١٢,٤)	$\frac{35 \times 57}{150}$ (١٣,٣)	صفر -
٣٦	$\frac{36 \times 40}{150}$ (٩,٦)	$\frac{36 \times 53}{150}$ (١٢,٧)	$\frac{36 \times 57}{150}$ (١٣,٧)	٥ -
٣٣	$\frac{33 \times 40}{150}$ (٨,٨)	$\frac{33 \times 53}{150}$ (١١,٧)	$\frac{33 \times 57}{150}$ (١٢,٥)	١٠ -

٣٢	$\frac{٣٢ \times ٤٠}{١٥٠}$ (٨, ٥)	$\frac{٣٢ \times ٥٣}{١٥٠}$ (١١, ٣)	$\frac{٣٢ \times ٥٧}{١٥٠}$ (١٢, ٢)	- ١٥
١٤	$\frac{١٤ \times ٤٠}{١٥٠}$ (٣, ٨)	$\frac{١٤ \times ٥٣}{١٥٠}$ (٤, ٩)	$\frac{١٤ \times ٥٧}{١٥٠}$ (٥, ٣)	٢٠ إلى أقل من ٢٥
١٥٠	٤٠	٥٣	٥٧	المجموع

ولاختبار فرض العدم في هذا المثال تطبق الصيغة التالية:

$$\text{مربع كاي} = \frac{\sum (n - \text{ق})^2}{\text{ق}}$$

$$\begin{aligned} & \frac{\sum (٩, ٣ - ٥)^2}{٩, ٣} + \frac{\sum (١٢, ٤ - ٩)^2}{١٢, ٤} + \frac{\sum (١٣, ٣ - ٢١)^2}{١٣, ٣} = \\ & \frac{\sum (١٢, ٥ - ٩)^2}{١٢, ٥} + \frac{\sum (١٢, ٧ - ١١)^2}{١٢, ٧} + \frac{\sum (١٣, ٧ - ١٨)^2}{١٣, ٧} + \\ & \frac{\sum (١٢, ٢ - ٨)^2}{١٢, ٢} + \frac{\sum (٨, ٨ - ٩)^2}{٨, ٨} + \frac{\sum (١١, ٧ - ١٥)^2}{١١, ٧} + \\ & \frac{\sum (٥, ٣ - ١)^2}{٥, ٣} + \frac{\sum (٨, ٥ - ١١)^2}{٨, ٥} + \frac{\sum (١١, ٣ - ١٣)^2}{١١, ٣} + \\ & \sum ١, ٤٤ = \frac{\sum (٣, ٨ - ٨)^2}{٣, ٨} + \frac{\sum (٤, ٩ - ٥)^2}{٤, ٩} + \end{aligned}$$

وحيث أن درجات الحرية = (د - ١) (ص - ١) = (١ - ٣) (١ - ٥)

$$\Lambda = ٤ \times ٢ =$$

ومستوى الدلالة المطلوب هو ٠,٠٥ ، فإن قيمة مربع كاي النظرية من توزيعها في الجدول هي ١٥,٥١ . وبما أن القيمة المحسوبة لمربع كاي (٢١,٤٤) أكبر من القيمة النظرية فإننا نرفض فرض العدم عند مستوى الدلالة ٠,٠٥ أى أننا نرفض القول بأنه لا توجد اختلافات جوهرية بين درجات الانحدار للأنواع الثلاثة من التركيب الصخري. أو بمعنى آخر فإننا نقبل الفرض البديل وهو أنه ليس من المحتمل أن يكون الاختلاف المشاهد بين درجات انحدار السفوح للأنواع الثلاثة من التركيب الصخري في المناطق الثلاث راجعاً للصدفة وحدها، بل هو اختلاف «حقيقي» وجوهري يبرز أهميته تأثير اختلاف نوع التركيب الصخري على درجات انحدار السفوح التي تكونت فوقه.

ثانياً: اختبار كولموجوروف - سميرنوف

Kolmogorov - Smirnov Test

(اختبار «D» ، D - test)

يشبه هذا الاختبار اختبار مربع كاي في أنه يستخدم لقياس مدى التطابق بين توزيعين أحدهما توزيع فعلي والآخر توزيع نظري (احتمالي). ولكن يفضل اختبار كولموجوروف - سميرنوف على مربع كاي لأنه أسهل في التطبيق، كما أن استخدامه لا يتطلب شروطاً خاصاً مثلما يتطلب تطبيق اختبار مربع كاي. وفي كثير من الدراسات الجغرافية يتطلب التحليل الكمي التعرف أولاً على حقيقة ما إذا كانت بيانات عينة الدراسة تنتمي مثلاً لمجتمع تتمثل فيه خصائص نوع معين من التوزيعات الاحتمالية المعروفة مثل توزيع بواسون، أو التوزيع المعتدل (الطبيعي). فلو كان لدينا بيانات عن تكرار ظاهرة ما على فترات معلومة فإنه يمكن أن نضع فرضاً إحصائياً بعشوائية حدوث أو تكرار هذه الظاهرة في تلك الفترة، ونقوم بحساب التوزيع النظري المتوقع فيما لو كانت هذه البيانات تتوزع حسب توزيع بواسون أو تتوزع توزيعاً معتدلاً. ثم يتم اختبار هل التوزيع الفعلي لبيانات هذه الظاهرة مطابق للتوزيع النظري المتوقع بحساب قيمة «D» (أى الفرق بين احتمالات كل من التوزيعين الفعلي والمتوقع). ودرجات الحرية لقيمة «D» تساوى مجموع التكرارات الكلية لبيانات الظاهرة.

وسوف نوضح خطوات حساب قيمة «د» لاختبار كولموجوروف - سميرنوف لبيانات جدول التوزيع التكرارى لهبوب العواصف الرعدية فى كل سنة لفترة ١٠٠ سنة.

وبما أن هذه الظاهرة تعد من الظواهر النادرة الحدوث، فإن توزيعها التكرارى المشاهد يتطابق، بدرجة كبيرة، مع توزيع بواسون الاحتمالى (النظرى). وبحساب قيمة «د» يمكن معرفة هل الاختلاف بين التوزيع النظرى أو المتوقع (توزيع بواسون) والتوزيع الفعلى أو المشاهد (الحقيقى) يرجع إلى الصدفة أم هو اختلاف حقيقى. وخطوات حساب قيمة «د» موضحة بالجدول رقم (١٣ - ٥).

جدول رقم (١٠ - ٥)

التوزيع الفعلى (المشاهد) لهبوب العواصف الرعدية
فى فترة ١٠٠ سنة والتوزيع النظرى (بواسون) لها

الاحتمال المتجمع			الاحتمال		عدد السنوات التي يتكرر فيها العواصف	عدد العواصف فى كل سنة
الفرق (د)	بواسون	المشاهد	بواسون	المشاهد		
٠,١٧	٠,١٣	٠,٣٠	٠,١٣	٠,٢٠	٣٠	صفر
٠,١٦	٠,٣٩	٠,٥٥	٠,٢٦	٠,٢٥	٢٥	١
٠,٠٤	٠,٦٦	٠,٧٠	٠,٢٧	٠,١٥	١٥	٢
٠,٠٣	٠,٨٥	٠,٨٢	٠,١٩	٠,١٢	١٢	٣
٠,٠٨	٠,٩٥	٠,٨٧	٠,٠١	٠,٠٥	٥	٤
٠,٠٦	٠,٩٩	٠,٩٣	٠,٠٤	٠,٠٦	٦	٥
٠,٠٥	١,٠٠	٠,٩٥	٠,٠١	٠,٠٢	٢	٦
٠,٠٤	١,٠٠	٠,٩٦	٠,٠٠	٠,٠١	١	٧
٠,٠٤	١,٠٠	٠,٩٦	٠,٠٠	٠,٠٠	٠	٨
٠,٠٤	١,٠٠	٠,٩٦	٠,٠٠	٠,٠٠	٠	٩
٠,٠٠	١,٠٠	١,٠٠	٠,٠٠	٠,٠٤	٤	١٠
			١,٠٠	١,٠٠	١٠٠	المجموع

ويلاحظ من الجدول السابق أننا قمنا بحساب الاحتمال المشاهد للتوزيع وذلك بتحويل التكرارات المطلقة إلى احتمالات أو تكرارات نسبية عن طريق قسمة كل تكرار من التكرارات التي تحدث بها العواصف على المجموع الكلى لعدد العواصف. ولاختبار فرض عشوائية التوزيع، فإننا نقارن هذا التوزيع الاحتمالى المشاهد بالتوزيع النظرى أو توزيع بواسون الاحتمالى الذى يناسب مثل هذه الحالة من المقارنة والتي تفترض أن حدوث العواصف الرعدية يتكرر عشوائياً فى الفترة الزمنية (١٠٠ سنة) قيد الفحص. وتجرى المقارنة لمعرفة مدى تطابق أو اقتراب الاحتمالات المشاهدة من الاحتمالات النظرية باختبار جودة التوفيق Goodness of Fit بينهما. وفى هذا المجال يمكن استخدام أو تطبيق إختبار كولموجوروف - سميير نوف لأداء هذه المهمة.

وعند مقارنة توزيعين احتماليين بواسطة إختبار كولموجوروف - سمييرنوف فإنه يجب تحويلهما إلى توزيعات احتمالية متجمعة Cumulative Probability distributions. ويكون فرض العدم لاختبار جودة التوفيق هو أن بيانات العينة التى ينتج عنها توزيعها احتمالياً فعلياً (مشاهداً) تكون لعينة مسحوبة من مجتمع يمتلك توزيعاً نظرياً خاصاً يشابه، فى هذه الحالة، توزيع بواسون الاحتمالى. وتحسب إحصائية الاختبار «د» على أساس أنها عبارة عن أقصى فرق (اختلاف) مطلق بين الاحتمال المتجمع لكل من التوزيعين الفعلى (المشاهد) والنظرى. ومن الجدول السابق نجد أن أقصى فرق (أو قيمة إحصائية الاختبار «د») هو ٠,١٧.

وحيث أن قيمة «د» المحسوبة (٠,١٧) أكبر من القيمة النظرية (٠,١٤) من الجدول الخاص بها (أنظر ملحق الجداول الإحصائية فى نهاية الكتاب) بدرجات الحرية ١٠٠ (مجموع التكرارات) وفى مستوى دلالة ٠,٠٥، فإننا نرفض فرض العدم عند مستوى الدلالة المعلوم. ويعنى هذا أن هناك احتمال أقل ٠,٠٥ (أو فى ٥ حالات فى كل ١٠٠ حالة) أن تمثل بيانات العينة توزيعاً له خصائص توزيع بواسون الاحتمالى. وتكون النتيجة النهائية بالنسبة للباحث الجغرافى هى أن التوزيع التكرارى الفعلى لحدوث العواصف الرعدية لايمكن أن يكون توزيعاً عشوائياً أو أن احتمال أن يكون كذلك ضعيف للغاية.

ثالثاً: اختبار مان - هويتى

The Mann - Whitney Test

(اختبار (U - test

يعتبر اختبار مان - هويتى من أبسط الأساليب الكمية غير الباراميتريّة التي تبحث في مقارنة مجموعتين من بيانات المعاينة لبيان ما إذا كانت هاتان المجموعتان مسحوبتين من مجتمعين مختلفين ولهما نفس المتوسط الحسابى أم مسحوبتين من مجتمع واحد، وهو بذلك يعد اختبار بديل لاختبار ستودنت - ت. ولا يشترط اختبار مان - هويتى، مثله في ذلك مثل بقية الأساليب اللاباراميتريّة الأخرى، أن يكون توزيع البيانات لكل عينة توزيعاً متماثلاً (معتدلاً)، كما لا يرتبط تطبيقية بنوع معين من أنواع البيانات ولكن يفضل استخدامه إذا كانت البيانات من نوع البيانات الترتيبية Ordinal (المزدوجة) والمتساوية وغير المتساوية في عدد مفرداتها.

ويستخدم اختبار مان - هويتى (U) لاختبار دلالة الفرق بين وسيطى عينتين أو اختبار فرض العدم القائل بأن العينتين مسحوبتان من مجتمع واحد وبالتالي يجب أن لا يكون هناك اختلافاً جوهرياً أو حقيقياً بين بيانات العينتين، وأن أى اختلاف بينهما إنما يرجع الصدفة. وتحسب قيمة الاختبار (U) من المعادلتين الآتيتين معاً:

$$U_1 = n_1 \times n_2 + \frac{n_1(n_1 + 1)}{2} - \text{محت } 1$$

$$U_2 = n_1 \times n_2 + \frac{n_2(n_2 + 1)}{2} - \text{محت } 2$$

حيث n_1 هي حجم العينة الأولى، n_2 هي حجم العينة الثانية، محت 1 هي مجموع الرتب للعينة الأولى، محت 2 هي مجموع الرتب للعينة الثانية.

وتوزيع مان - هويتى غير منتظم ويقترّب من التوزيع المعتدل كلما زادت أحجام العينات، لذلك في الحالات التي يزداد فيها حجم العينات فإنه يمكن

استخدام توزيع (ز) - جداول المساحات تحت المنحنى المعتدل - للاستدلال على
الفرض الموضوع للاختبار.

وسنوضح فى الأمثلة التالية خطوات حساب الفرق بين بيانات عينتين بواسطة
اختبار مان-هويتنى (ى).

مثال (١):

لدراسة نسبة المحتوى المائى (نسبة الرطوبة) فى التربة أخذت عدة قياسات فى
منطقتين زراعتين فكانت نسبة الرطوبة فى التربة فى كل منها كما يلى:

المنطقة الأولى: ١٥,٢ ١٠,٧ ٨,٦ ٩,٣ ١٢,٤ ١١,١

المنطقة الثانية: ٨,٤ ٩,٣ ٨,٧ ١٠,٢ ١٠,٠

وفرض العدم فى هذه الحالة هو أن نسبة المحتوى المائى (نسبة الرطوبة) فى
المنطقة الأولى أكبر من أو تساوى مثلتها فى المنطقة الثانية بمستوى دلالة ٠,٠٥
ولاختبار الفرض السابق نحسب قيمة (ى) وذلك بافتراض أن صفة نسبة المحتوى
المائى للتربة لا تتوزع توزيعاً معتدلاً، كما أن كل عينة مستقلة عن الأخرى وكلاً
العينتين أخذت بطريقة عشوائية. وخطوات الحساب موضحة فى الجدول التالى:

جدول رقم (١٠-٦)

حساب قيمة (ى) من بيانات عيتين لنسبة المحتوى المائى فى التربة

العينة الأولى	الرتبة	العينة الثانية	الرتبة
١٥,٢	١٢	٨,٤	١
١٠,٧	٨	٩,٣	٤,٥
٨,٦	٢	٨,٧	٣
٩,٣	٤,٥	١٠,٢	٧
١٢,٤	١٠	١٠,٠	٦
١١,١	٩		
مجم	٤٤,٥	٢١,٥	٣٥ = ٥

من الجدول السابق يتضح أننا قمنا بترتيب جميع المفردات للمعينتين في ترتيب واحد، كما يلاحظ أننا أعطينا رتبة موحدة للمفردات المتساوية القيمة فمثلاً المفردة ٩,٣ تكررت مرتين في البيانات، وبما أن موضعها يحتل الرتبة الرابعة والخامسة بين الترتيب الكلى، فإننا أعطينا لها الترتيب ٤,٥ (أى $(٤ + ٥) \div ٢$). وبالمثل إذا كان لدينا ثلاث مفردات متطابقة في قيمها تحتل مواقعها الرتبة الخامسة والسادسة والسابعة فإننا نعطي لكل منها الرتبة ٦ (أى $(٥ + ٦ + ٧) \div ٣$)، وهكذا. ولاختبار الفرق بين العينتين في نسبة المحتوى المائى في التربة نحسب قيمة (ى) باستخدام كل من المعادلتين السابقتين على النحو التالى:

$$٤٤,٥ - ٢١ + ٣٠ = ٤٤,٥ - \frac{(١ + ٦) ٦}{٢} + ٥ \times ٦ = (ى)$$

$$٦,٥ = ٤٤,٥ - ٥١ =$$

$$٢١,٥ - ١٥ + ٣٠ = ٢١,٥ - \frac{(١ + ٥) ٥}{٢} + ٥ \times ٦ = (ى)$$

$$٢٣,٥ = ٢١,٥ - ٤٥ =$$

ولاختبار معنوية قيمة (ى) نحتاج دائماً إلى أصغر قيمة من قيمتى (ى) المحسوبتين من المعادلتين السابقتين. ولتسهيل عملية الحساب فإنه يمكن الحصول على قيمة (ى) من أية معادلة من المعادلتين ومنها يمكن الحصول على القيمة الأخرى وذلك لأن مجموع القيمتين لا بد وأن يساوى حاصل ضرب حجم العينة الأولى في حجم العينة الثانية (١٠×٢) . ففي المثال $١٠ \times ٢ = ٣٠$ ، فإذا كانت قيمة (ى) الأولى تساوى ٦,٥ فإن قيمة (ى) الأخرى تساوى ٢٣,٥ (أى $٦,٥ - ٣٠$).

ونظراً لأن الفرض البديل في هذه الحالة محدد الاتجاه فإننا نختار إختبار المعنوية من طرف واحد One - tailed test أو رفض فرض العدم بمستوى الدلالة المطلوب. وتكون النتيجة أنه تبعاً لأن قيمة (ى) الصغرى المحسوبة وهى ٦,٥ أكبر من قيمة (ى) النظرية التى تساوى ٥ عند مستوى الدلالة ٠,٥، لاختبار الطرف الواحد وعندما تكون $١ = ٦$ ، $٢ = ٥$ (انظر ملحق الجداول الاحصائية)

فإننا نقبل فرض العدم، أى أن نسبة المحتوى المائى فى التربة (نسبة الرطوبة) فى المنطقة الأولى تساوى أو أكبر «جوهرياً» من مثلتها فى المنطقة الثانية.

وتجدر الإشارة هنا إلى أن اختبار الدلالة (المعنوية) الخاص باختبار مان-هويتنى (ى)، دون معظم اختبارات الدلالة الأخرى، يرفض فيه فرض العدم الموضوع للاختبار عند مستوى دلالة معين إذا كانت قيمة (ى) المحسوبة أقل من أو تساوى قيمة (ى) النظرية (القيمة الحرجة U-Critical Value).

مثال (٢):

فى دراسة لمعرفة اختلاف تأثير العوامل البحرية والنهرية على استدارة الرواسب الحصوية جمعت بيانات عن معامل الاستدارة^(١) من عينتين عشوائيتين من الحصى: تمثل إحداهما مجتمع الرواسب على الأرصفة البحرية (الشواطئ المرتفعة) والأخرى تمثل مجتمع الرواسب الدلتاوية النهرية، فكانت معاملات استدارة الحصى فى كل عينة كما هو موضح بالجدول التالى؛ والمطلوب اختبار الفرض القائل بوجود اختلاف فى تأثير كل من العوامل البحرية والنهرية على استدارة الرواسب. الحصوية أو بعبارة أخرى هل يمكن الحكم على أن العينتين مسحورتان من مجتمع واحد بمستوى دلالة ٠,٠٥.

(١) معامل الاستدارة هو مقياس كمى تقاس به استدارة الرواسب الحصوية، ويمكن حسابه من المعادلة:

$$\frac{2}{L} \times 1000 \text{ حيث } L \text{ هى قيمة نصف قطر أصغر تحدب للحصوة، } L \text{ هى أكبر طول للحصوة}$$

وتراوح معامل الاستدارة للرواسب الحصوية ١ - ١٠٠٠ (الكرة المثالية)

جدول رقم (١٠ - ٧)

حساب قيمة (ى) من بيانات عينتين لاستدارة الرواسب الحصوية

الرتبة	معامل استدارة الحصى النهري	الرتبة	معامل استدارة الحصى البحرى
٦	٣٢٥	١٢	٣٤٩
٢٣	٤٨٥	١	٢٥٨
٢٢	٤٥٦	٧	٣٣٥
٢٠	٤٣٨	٤	٣١٧
٢١	٤٥٥	٣	٣١٠
٨	٣٨٨	٢	٣٠٤
١٦	٣٦٩	١٧	٣٧١
١٨	٣٨٩	٥	٣٤٠
١٥	٣٦٥	١٠	٣٤٤
١٤	٣٥١	١١	٣٤٥
١٣	٣٥٠	١٩	٤٠٥
		٥	٣٠٤
المجموع (محت ٢) = ١٧٦		المجموع (محت ١) = ١٠٠	
٢ = ١١		١٢ = ١٢	

بعد إعطاء قيم معاملات الاستدارة فى العينتين الرتب الخاصة بهما، وبعد جمع رتب كل عينة على حدة، نقوم بحساب قيمة (ى) على النحو التالى:

$$\text{قيمة (ى) المحسوبة للعينة الأولى} = \text{ن}_١ \times \text{ن}_٢ + \frac{\text{ن}_١ (\text{ن}_١ + ١)}{٢} - \text{محت ١}$$

$$= ١٢ \times ١١ + \frac{١٢ (١٢ + ١)}{٢} - ١٠٠ =$$

$$= ١٣٢ + ٧٨ - ١٠٠ = ١١٠$$

$$\text{قيمة (ى) المحسوبة للعينة الثانية} = \text{ن}_١ \times \text{ن}_٢ + \frac{\text{ن}_٢ (\text{ن}_٢ + ١)}{٢} - \text{محت ٢}$$

$$= ١٢ \times ١١ + \frac{١١ (١١ + ١)}{٢} - ١٠٠ =$$

$$= ١٣٢ + ٧٨ - ١٠٠ = ١١٠$$

وبطبيعة الحال فإننا نحتاج إلى القيمة الأصغر من قيمتي (ى) المحسوبيتين وهى (٢٢) للاستدلال عن معنوية الفرض الموضوع للاختبار. وبمقارنة قيمة (ى) المحسوبة (٢٢) بقيمة (ى) النظرية فى جدول الاختبار من طرف واحد فى المستوى دلالة ٠,٠٥، ولحجم العينة ١٢، ١١ وهى (٣٨) فإنه يمكن فى هذه الحالة رفض فرض العدم القائل أن العينتين مسحوبتان من مجتمع واحد، أى أن هناك تأثير للعوامل البحرية والنهرية، كل على حدة، على استدارة الرواسب الحصوية بنوعيتها - يظهر فى وجود فروق جوهرية بين معامل استدارة الرواسب الحصوية فى العينتين.

مثال (٣):

إذا كانت لدينا عينتين عشوائيتين تمثل كل منهما إنتاج فول الصويا (طن / فدان) لعدد من الأحواض الزراعية فى منطقتين وذلك بفرض اختبار أن كل من متوسطى العينتين متساويين أو بمعنى آخر هل العينتان مأخوذتان من مجتمعين مختلفين أم من مجتمع واحد؟ وفى حالة عدم تساوى المتوسطين فإننا نتوقع أن متوسط العينة الأولى يزيد عن متوسط العينة الثانية، أى أن الاختبار المناسب هو اختبار (ى) فى مجموعتين من طرف واحد بمستوى دلالة ٠,٠٥ وكانت بيانات العينتين موضحة فى الجدول التالى:

جدول رقم (١٠ - ٨)

إنتاج فول الصويا (طن/ فدان) في منطقتين مرتبة حسب مقدار الإنتاج

الرببة (ت)	الإنتاج في المنطقة الثانية	الرببة (ت)	الإنتاج في المنطقة الأولى
٦	٣,٠٥	١٣	٣,٢٨
٢	٢,٨٦	١١	٣,١٩
٩	٣,١٦	٢٣	٣,٦٣
١٠	٣,١٧	٢٠	٣,٥٦
١٢	٣,٢٠	٢١	٣,٦١
١٥	٣,٣٣	٤	٢,٩٨
٣	٢,٩٦	١٨	٣,٤٥
١٤	٣,٢٩	١٧	٣,٤٠
٨	٣,١٥	١٩	٣,٤٩
٥	٣,٠١	١٦	٣,٣٤
٢٢	٣,٦٤		
٧	٣,١٤		
١	٢,٨٢		
١١٤	١٣ = ٣	١٦٢	المجموع = ١٠ = ١

لاختبار الفرض السابق نحسب قيمة (ي) بالمعادلتين السابقتين كما يلي:

$$(ي) = ١٠ \times ١٣ + \frac{١٠(١ + ١٣)}{٢} - \text{محت}$$

$$= ١٦٢ - \frac{١٠(١ + ١٣)}{٢} + ١١ \times ١٢ =$$

$$= ٢٣ = ١٦٢ - ٥٥ + ١٣٠ =$$

$$(ي) = ١٠ \times ١٣ + \frac{١٠(١ + ١٣)}{٢} - \text{محت}$$

$$114 - \frac{(1 + 13) 13}{2} + 11 \times 12 =$$

$$107 = 114 - 91 + 130 =$$

وحيث أن قيمة (ى) الصغرى المحسوبة وهى (٢٣) أقل من قيمة (ى) النظرية لحجم العينة ١٠، ١٣ بمستوى دلالة ٠,٠٥ وفى اختبار الطرف الواحد وهى (٣٧) فإننا نرفض فرض العدم ونقبل الفرض البديل وهو أن متوسط إنتاج فول الصويا فى الأحواض الزراعية فى المنطقة الأولى (العينة الأولى) يزيد عن متوسط الإنتاج للأحواض الزراعية فى المنطقة الثانية (العينة الثانية) بمستوى دلالة ٠,٠٥ والنتيجة التى يمكن أن نستخلصها، اعتماداً على البيانات السابقة تؤكد تأثير العوامل الجغرافية فى الإنتاج الزراعى بعمامة وإنتاج فول الصويا بخاصة فى المنطقتين.

رابعاً: اختبار ويلكوكسون

Wilcoxon Test

(اختبار وق)

يعد اختبار ويلكوكسون من أبسط أساليب المقارنة الاحصائية اللاباراميتريية التى يمكن استخدامها لاختبار دلالة الفروق (الاختلافات) بين رتب أزواج القيم (المفردات) المتماثلة للمتغير موضع الاختبار. ويتطلب الاختبار إذن أن تكون البيانات الأصلية من نوع بيانات الفترة ولكن جدولتها تكون على شكل رتب، كما يعتمد مستوى الدلالة (المعنوية) المستخدم لرفض فرض العدم لهذا الاختبار على اتجاه ومقدار الاختلاف بين الرتب لكل زوج من أزواج القيم المتماثلة (أى عندما تكون إحدى القيم (أ) أكبر من القيمة الأخرى (ب) أو العكس). كما يشترط عند استخدام اختبار ويلكوكسون لأزواج المفردات المتماثلة فى العدد أن تكون مفردات كل زوج متجانسة وتخصص إحدى المفردات من كل زوج للعينة الأولى بينما تخصص المفردة الأخرى للعينة الثانية.

ولا يقتصر تطبيق إختبار ويلكوكسون على الحالات التى يكون فيها حجم العينة صغيراً (ن أقل من ٣٠) فحسب، بل أنه يستخدم كذلك فى الحالات التى يكون فيها حجم العينة كبيراً (ن أكبر من ٣٠)، ولا تختلف عملية الاختبار فى الحالتين إلا فى الخطرات الحسابية الأخيرة. ويتلخص الاختبار فى كل حالة فى اعتبار الفروق بين رتبتي كل زوج من القيم عينة عشوائية من مجتمع تمثل جميع الفروق الممكنة، واختبار هل العيتان مأخوذتان من مجتمع واحد أم من مجتمعين لهما نفس المتوسط الحسابي أو مسحوبتان من مجتمعين مختلفين. ولتوضيح طريقة مقارنة بيانات العينات الصغيرة الحجم (ن أقل من ٣٠) نذكر المثال التالي:

مثال:

لمقارنة درجات الانحدار على جانبي أحد الأودية النهرية (باتجاه شمالي غربي «ش.غ» لأحد الجوانب واتجاه جنوبي شرقي «ج.ق» للجانب الآخر) أخذت القياسات الموضحة فى الجدول التالي (جدول رقم ١٠ - ٩). وتمثل القياسات المزدوجة والمتماثلة فى العدد بالجدول المسافة الكنتورية (بالمليمتر) المأخوذة من قطاعات عرضية ذات أبعاد متساوية فيما بينها على خريطة كنتورية بمقياس كبير. وكما نعلم أن هناك علاقة وطيدة بين المسافة الكنتورية ودرجة الانحدار، فكما اتسعت المسافة بين خطوط الكنتور كلما دل ذلك على ضعف انحدار سطح الأرض والعكس صحيح. والمطلوب هو إختبار ما إذا كان الفرق بين متوسطي درجات الانحدار لجانبي الوادي يمثل فرقاً جوهرياً (حقيقياً) أو أنه فرق يرجع إلى الصدفة المطلقة نتيجة خطأ المعاينة. وبعبارة أخرى فإن فرض العدم المطلوب اختباره هو أن مقدار الانحدار لا يختلف على منحدرات جانبي الوادي فى مقابل الفرض البديل أن مقدار الانحدار يختلف ويتنوع على جانبي الوادي من مكان لآخر حسب درجة التوجيه Aspect وذلك بمستوى دلالة ٠,٠٥.

جدول رقم (١٠ - ٩)

طريقة حساب إختبار ويلكوكسون من بيانات الفترة الكتتورية
(بالمليمتر) على جانبي أحد الأودية النهرية

الرتب الخاصة		رتب الفرق	الفرق أ - ب	الجانِب المواجه للجنوب الشرقى (ب)	الجانِب المواجه للشمال الغربى (أ)
أ > ب	أ < ب				
١		١	١	١١	١٠
		١	صفر	٧	٧
٦		٦	٤	١٣	٩
	٢,٥	٢,٥	٢	٨	١٠
	٤,٥	٤,٥	٣	١٢	١٥
٤,٥		٤,٥	٣	١٥	١٢
٢,٥		٢,٥	٢	١٠	٨
٧		٧	٦	١٥	٩
٢١ = ٣	٧ = ٣				المجموع

١- الفروض: H_0 : متوسط انحدار السفوح الشمالية الغربية = متوسط انحدار السفوح الجنوبية الشرقية.

H_1 : متوسط انحدار السفوح الشمالية الغربية \neq متوسط انحدار السفوح الجنوبية الشرقية.

٢- مستوى المعنوية $(\alpha) = ٠,٠٥$ والاختبار فى هذه الحالة من طرفين تبعاً لافتراض تساوى المتوسطين، ولعدم تحديد أى من الجانبين أشد انحداراً من الآخر. وبذلك إذا كان احتمال حدوث هذه البيانات بالصدفة تحت شروط الاختبار أقل من ٢٠٠١ فإن نرفض فرض العدم، أى أن الفرضية تكون فى هذه الحالة غير صحيحة.

٣- بعد حساب الفرق بين أزواج القيم (مع إهمال الإشارة) تعطى الفرق رتباً تختلف حسب مقدار الفرق. ويعطى للفرق المتساوية متوسط الرتب الطبيعية لهذه الفرق. فمثلاً الفرق ٢، ٢ رتبتهما على الترتيب هما ٢، ٣ ومتوسطهما هو ٢، ٥، وبالمثل الفرق ٣، ٣ رتبتهما على الترتيب هما ٤، ٥ وتكون الرتبة المتوسطة لهما هي ٤، ٥. ثم نفصل بين ترتيب (رتب) الفرق المحسوبة بين العينتين الأولى (أ) و (ب)، أى عندما تكون قيم العينة الأولى أكبر من القيم فى العينة الثانية (ب)، والعكس عندما تكون قيم العينة الثانية (ب) أكبر من العينة الأولى (أ). ويكون أقل مجموع لهذه الرتب هو القيمة المختبرة، والتي يرمز لها بالرمز (ق)، وذلك على أساس أنه إذا كان مجموع الرتب لكل من r_1 ، r_2 متساوياً (أى أن $r_1 - r_2 = \text{صفر}$) فإن فرض العدم يكون صحيحاً طالما أن الفرق بين أزواج القيم يتوزع عشوائياً بين القطاعات التي تكون فيها مفردات العينة (أ) أكبر من مفردات العينة (ب) وبالمثل عندما تكون (ب) أكبر من (أ). أما إذا كان الفرق بين كل من r_1 ، r_2 أكبر من الصفر فإن ذلك يتخذ دليلاً كلما دل ذلك على عدم صحة فرض العدم. وحيث أن مجموع الرتب الخاصة ($r_1 + r_2$) يكون ثابتاً لعينة حجمها (ن) فإن هذا المجموع يمكن وضعه فى الصيغة التالية:

$$\frac{1}{n} (n + 1)$$

وبذلك فإنه كلما زاد الفرق بين r_1 ، r_2 كلما قلت القيمة الصغرى لهما وكلما دل ذلك على عدم صحة فرض العدم وبالتالي يمكن رفضه.

٤- منطقة الرفض: بالرجوع إلى جدول اختبار ويلكوكسون (راجع ملحق الجداول الاحصائية) والتي تتضمن القيم الحرجة (ق) للعينات المختلفة الحجم: من $n = 6$ حتى $n = 25$ (إذ لا يمكن تطبيق اختبار ويلكوكسون على العينات ذات الحجم أقل من ٦ أزواج من المفردات يمكن استخراج القيمة الحرجة التى نقارن بها القيمة المحسوبة (ق). وحيث أن مستوى الدلالة المطلوب هو ٠,٠٥ وأن الفرض البديل غير محدد الاتجاه فإن الاختبار للمثال بين أيدينا من نوع الاختبار من الطرفين. وطبقاً للأسس الموضوعية سابقاً فإذا كانت قيمة (ق) المحسوبة أقل من

القيمة الحرجة المناظرة لها بالنسبة لحجم العينة قيد الاختبار فإننا نرفض فرض العدم ونستنتج أن الفرق بين العيتين هو فرق جوهري أو حقيقى عند مستوى الدلالة ٠,٠٥ .

٥- الاستنتاج: وبما أن القيمة الحرجة لهذا المثال هي ٢ عند $n = ٧$ (بعد أن أسقطنا من حسابنا أحد أزواج القيم من الحجم الأصلي للعينة لعدم وجود فرق بين قيمتيه) والقيمة المحسوبة (ق) هي ٧، فإن بيانات العينة لا تقدم دليلاً كافياً لرفض فرض العدم. أو بعبارة أخرى أنه تحت شروط الاختبار نجد أن قيمة (ق) المحسوبة أكبر من القيمة الحرجة بمستوى معنوية ٠,٠٥، وبالتالي نقبل فرض العدم ونستنتج أن هذا الفرق يرجع إلى الصدفة المطلقة. ويعنى قبولنا لفرض العدم في هذه الحالة أنه لا توجد اختلافات جوهريّة بين درجات الانحدار على جانبي هذا الوادى النهري تبعاً لاختلاف درجة التوجيه Aspect، مما يقلل من دور هذا العامل البيئي في التأثير على درجات الانحدار للمنحدرات في حوض هذا النهر.

وفى حالة العينات الكبيرة (ن أكبر من ٣٠) يصمم اختبار ويلكوكسون بنفس الطريقة التي اتبعناها سابقاً لاختبار الفروض التي تتعلق بالمقارنة بين المتوسطات لعينات صغيرة أقل من ٣٠ من أزواج القيم المتماثلة فى العدد، فيما عدا أن الاستنتاج (أى قرار قبول أو رفض فرض العدم) فى هذه الحالة يختلف قليلاً عن مثيله للعينات الصغيرة على أساس أن الفرق بين الرتب يدخل فى الاعتبار عند تقرير ما إذا كان صغر قيمة (ق) كافياً لتأكيد رفض فرض العدم. ونظراً لأن جدول القيم الحرجة لاختبار ويلكوكسون لا يشمل على قيم حرجة لمجموع الفرق بين أكثر من ٢٥ زوجاً من قيم العينات، فإنه فى حالة العينات الكبيرة يؤول توزيع القيمة (ق) تحت شروط الاختبار إلى الطرف الأيسر لمنحنى التوزيع المعتدل الذى متوسطه الحسابى يساوى نصف المجموع الكلى للرتب [أى يساوى $(١+n)/٢$]،

وانحرافه المعيارى يساوى $\sqrt{\frac{[n(n+1)(2n+1)]}{24}}$. وبالتالى فإن احتمال

(ح) بأن تكون القيمة (ق) - تحت شروط الاختبار وفرض العدم - أقل من أو تساوى القيمة المحسوبة يمكن الحصول عليه بواسطة:

أ- تحويل الإحصائية (ق) إلى قيمة معيارية من قيم الإحصائية (ز) Z - Score: عن طريقة إيجاد الفرق بينهما وبين متوسط توزيع المعاينة وقسمة هذا الفرق على الانحراف المعياري لتوزيع المعاينة، أى أن:

$$Z = \frac{Q - \frac{1}{2}(n+1)}{\sqrt{\frac{[n(n+1)(2n+1)]}{24}}}$$

ب- إيجاد قيمة الاحتمال (ح) المقابلة للقيمة المحسوبة من جداول التوزيع المعتدل المعيارى حسب نوع الاختبار (من طرف واحد أو من طرفين)، فإذا كانت قيمة الاحتمال المحسوبة أقل من قيمة احتمال مستوى الدلالة (α) فإننا نرفض فرض العدم ونقبل الفرض البديل، والعكس إذا كان احتمال القيمة المعيارية أكبر من احتمال مستوى الدلالة (α) فإن فرض العدم لا يمكن رفضه.

ولتوضيح استخدام اختبار ويلكوكسون فى حالة العينات الكبيرة نذكر المثال

التالى:

مثال:

الجدول رقم (١٠ - ١٠) يشتمل على بيانات لأعداد القوى العاملة (بالمئات) بمجموعة من القرى فى عام ١٩٧٠، ١٩٧٦. والمطلوب عن طريق استخدام اختبار ويلكوكسون تحديد احتمال أن مثل هذا الفرق الكبير بين مفردات العينتين، كما يبدو فى الجدول، يمكن حدوثه حتى إذا كانت مجتمعاتهما النظرية متشابهة ومتساوية فى معالمها.

(أ) = العمالة في ١٩٧٠ ، ب = العمالة في ١٩٧٦

وتكون خطوات الاختبار كما يلي:

١- H_0 : لا يوجد اختلاف بين أعداد القوى العاملة لعامي ١٩٧٠ و ١٩٧٦.

H_1 : يوجد اختلاف بين أعداد القوى العاملة لعامي ١٩٧٠ و ١٩٧٦

٢- مستوى الدلالة (المعنوية) $\alpha = ٠,٠٥$

٣- باتباع نفس الإجراءات السابقة لاختبار العينات الصغيرة نحسب قيمة (ق) من واقع البيانات المشاهدة، وهي لهذا المثال، تساوى ٨٥,٥.

٤- منطقة الرفض: تحول قيمة (ق) المحسوبة إلى قيمة معيارية من قيم (ز) المعيارية كما يلي:

$$z = \frac{Q - \frac{1}{2}(n+1)}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} = \frac{147}{49} = \frac{31 \times 30 \times \frac{1}{2} - 85,5}{\sqrt{\frac{61 \times 31 \times 30}{24}}}$$

(ب) نرجع إلى جدول التوزيع المعتدل المعياري للحصول على قيمة (ز) النظرية. ونظراً لأن الفرض البديل غير محدد الاتجاه فإن الاختبار يكون من طرفين، وبالتالي فإن الاحتمال المقابل للقيمة المحسوبة (ز = ٣) هو ٠,٠٠٣.

٥- الاستنتاج: تحت شروط الاختبار نجد أن احتمال قيمة (ز) المحسوبة من القيمة (ق) وهو ٠,٠٠٣، وأقل بكثير من مستوى الرفض ٠,٠٥، لذا لا يمكن أن يرجع هذا الفرق للصدف المطلقة، ونرفض فرض العدم القائل بأنه لا يوجد فرق معنوي بين أعداد القوى العاملة في عام ١٩٧٠ و ١٩٧٦، أى أن البيانات تمثل إختلافاً حقيقياً في عدد القوى العاملة بين الستين. وأكثر ما يمكن تفسيره جغرافياً من

هذا المثال هو أنه قد يوضح أن الاختلاف في أعداد القوى العاملة في العينتين (١٩٧٠، ١٩٧٦) قد يرجع إلى عامل زيادة الحركة في مجتمع القوى العاملة أو عامل انتشار الصناعة في المناطق الريفية أو إلى العاملين معاً. ولكن لا يمكن تأكيد هذه العوامل إلا عن طريق البحث المستفيض والدراسة التفصيلية وذلك لأن الأساليب الكمية كما نعلم، تثير كثيراً من التساؤلات أكثر من الإجابة عليها.

خامساً: اختبار كروسكال - واليس

The Kruskal - Wallis Test

(اختبار «هـ» H - test)

يستخدم اختبار كروسكال - واليس في حالة المقارنة وبيان مدى التجانس أو الاختلاف بين ثلاث عينات أو أكثر ومدى صلتها بالمجتمع الأصلي الذي تمثله. ولذا فإنه يعتبر اختباراً يصلح كبديل لتحليل التباين، طالما أنه لا يشترط، مثل غيره من أساليب المقارنة اللاباراميتريّة، أن يكون توزيع مفردات أو قيم العينات متصفاً بصفة الاعتدالية، بل يتطلب فقط أن تكون البيانات من نوع البيانات الترتيبية (الرتب). وعلى الرغم من بساطة الاختبار وسهولة حسابه إلا أنه كأداة للتحليل لم يحظ بالاهتمام والاستخدام، حتى وقت قريب، في مجال البحث الجغرافي.

ويعتمد تحليل الاختلافات بين العينات باختبار «هـ» على فرض العدم القائل بأن العينات قد أخذت من مجتمعات لها توزيعات متطابقة، وأن أي اختلاف فيما بينها إنما يرجع إلى الصدفة المطلقة التي تتضمنها عملية المعاينة العشوائية. ويكون الفرض البديل المقابل لفرض العدم في هذه الحالة هو أن الاختلاف بين العينات يعكس الاختلافات بين توزيع المجتمعات التي سحبت منها. ويمكن اختبار الاحصائية «هـ» بحسابها من المعادلة الآتية:

$$H = \frac{12}{n(n+1)} \times \text{مجم} - \frac{3}{n} (1 + n)$$

حيث n هي المجموع الكلي للمفردات في كل العينات، r هي مجموع الرتب لكل عينة، n هي عدد المفردات في كل عينة على حدة، $\frac{r^2}{n}$ هو المجموع الكلي لمربعات مجموع الرتب المقسومة على عدد مفردات كل عينة على حدة. وهناك جداول محسوبة (راجع ملحق الجداول الإحصائية) لقيم «هـ» باحتمالات مختلفة لعينات ثلاث فقط بكل منها عدد من المفردات يتراوح بين مفردة واحدة وخمس مفردات. أما إذا زاد عدد المفردات عن خمس مفردات فإن قيمة «هـ» يكون لها توزيع احتمالي يتطابق مع التوزيع الاحتمالي لمربع كاي، بدرجات الحرية ($n - 1$) حيث n هي عدد العينات. ولذا لا يقتصر تطبيق اختبار «هـ» على العينات الصغيرة فقط ولكن يمكن أيضاً استخدامه لتحليل بيانات العينات الكبيرة. ونقصد بالعينات الصغيرة في هذا المقام بأنها العينات التي لا يزيد عددها عن ثلاث عينات بكل منها عدد من المفردات (أو القياسات «ن») لا يزيد عن 5 مفردات.

والمثال التالي يوضح خطوات حساب الاختبار للعينات الصغيرة (ق = 3، $n > 5$).

مثال (٢):

الجدول التالي (جدول رقم: ١٠ - ١١) يوضح درجات الانحدار للسفوح الناجمة عن فعل الحركة البطيئة للمواد التي تتألف رواسبها من المفتتات الصخرية غير المغطاة بغطاءات نباتية Unvegetated Scree slopes التي تفككت من ثلاثة أنواع مختلفة من التكوينات الصخرية.

جدول رقم (١٠ - ١١)

درجات الانحدار لسفوح المفتتات الصخرية حسب نوع الصخور التي تفككت منها

حجر جيري		شمس		حجر جيري	
الرتبة	درجة الانحدار	الرتبة	درجة الانحدار	الرتبة	درجة الانحدار
١٢	٣٠	٥	٢١	١	١٧
١٠	٢٦	٢	١٨	٤	٢٠
٦	٢٢	٩	٢٥	٣	١٩
		١١	٢٨	٨	٢٤
		٠٠	٠٠	٧	٢٣
ن = ٣ = ٣		ن = ٢ = ٤		المجموع ن = ١ = ٥	
٢٨ = ٣		٢٧ = ٢		٢٣ = ١	

١- الفروض: H_0 : لا يوجد اختلاف بين درجات الانحدار لسفوح المفتتات التي تكونت من أنواع مختلفة من التكوينات الصخرية، أو بعبارة أخرى أن هناك احتمالاً كبيراً أن ترجع الاختلافات المشاهدة في درجات الانحدار بين العينات الثلاث إلى الصدفة.

H_1 : أن الاختلافات في درجة الانحدار بين العينات تعكس الاختلافات بين المجتمعات التي تمثلها هذه العينات، أي أنها إختلافات جوهريّة حقيقية لا ترجع إلى الصدفة.

٢- مستوى الدلالة $(\alpha) = ٠,٠٥$.

٣- يمكن وضع الاختبار على الصورة الآتية:

$$H = \frac{12}{n(n+1)} \times \text{محرّف} \frac{r}{n} - 3(n+1)$$

٤- هذه القيمة لها توزيع احتمالي هـ حسب عدد المفردات في كل عينة، أى:

$$١٠ = ٥, \quad ٣ = ٤, \quad ٣ = ٣$$

٥- منطقة الرفض: على أساس الشروط السابقة من (١) إلى (٢) نجد أن القيمة الحرجة لاختبار هـ، التى تحدد الرفض هى ٥, ٦٣١ ونقبل فرض العدم إذا كانت قيمة هـ المحسوبة أقل من ٥, ٦٣١ أو لتكون قيمة هـ المحسوبة لها دلالة إحصائية لابد أن تساوى أو تزيد عن القيمة النظرية المقابلة لها.

٦- حساب قيمة هـ من واقع البيانات المشاهد، حيث:

$$\begin{array}{ll} ١٧ = ٢, & ٢٣ = ١, \\ ١٢ = ٣, & ٥ = ١, \\ ٢٨ = ٣, & ٣ = ٣, \\ ٣ = ٣, & \end{array}$$

$$(١ + ١٢) ٣ - \left(\frac{٢(٢٨)}{٤} + \frac{٢(٢٧)}{٤} + \frac{٢(٢٣)}{٥} \right) \times \frac{١٢}{(١ + ١٢) ١٢} = (هـ)$$

$$٢٩ - (٢٦١, ٣٣ + ١٨٢, ٢٥ + ١٠٥, ٨٠) \times , ٠٧٦٩ =$$

$$٣٩ - (٥٤٩, ٣٨٣٣ \times , ٠٧٦٩) =$$

$$٣, ٢٤٧ =$$

٧- الاستنتاج قيمة هـ المحسوبة من البيانات المشاهدة تقل عن القيمة الحرجة (٥, ٦٣١)، ولذلك لا يمكن رفض فرض العدم القائل بأنه لا يوجد فرق جوهري بين درجات الانحدار لسفوح المفتتات الصخرية التى تفككت من أنواع مختلفة من التكوينات الصخرية وذلك بمستوى دلالة ٠, ٠٥.

ويتبع نفس الأسلوب السابق عند تحليل بيانات العينات الكبيرة (ن أقل من ٣ أو تساوى ٣، ن أكبر من ٥) كما يتضح من المثالين التاليين.

مثال (٢):

الجدول التالى (جدول رقم ١٠ - ١٢) يشتمل على بيانات لعينة من الزوار ثلاث من الحدائق العامة فى أحد الأيام ممثلة فى المسافات (بالكيلو مترات) التى

قطعها زوار كل حديقة. والمطلوب اختبار فرض العدم القائل بأن العينات الثلاث للزوار تمثل مجتمعا واحدا للزوار أو مجتمعات متطابقة في مقابل الفرض البديل القائل بأن هذه العينات تمثل مجتمعات مختلفة من الزوار وذلك بمستوى دلالة ٠,٠٥.

جدول رقم (١٠ - ١٢)
طريقة حساب اختبار (H) للمسافات (بالكيلومتر) التي قطعتها ثلاث عينات من الزوار لثلاث من الحدائق العامة

الحديقة (أ)		الحديقة (ب)		الحديقة (ج)	
المسافة	الرتبة	المسافة	الرتبة	المسافة	الرتبة
٢٣	٧	١١	٣	١١٠	١٨
١٥	٤	١٧	٥	٨٤	١٦
٤٢	١٢	٣١	١١	٨٥	١٧
٨	١	٢٧	٩,٥	٤٥	١٣
١٠	٢	٦٣	١٤	٦٤	١٥
١٨	٦	٢٧	٩,٥	٢٥	٨
ن = ١٢		ن = ٦		ن = ٦	
٣٢ = ∑		٥٢ = ∑		٨٧ = ∑	

(١) الفروض H_0 : لا يوجد اختلاف بين المسافات للعينات الثلاث، أو بعبارة أخرى أن هذه العينات تمثل نفس المجتمع أو أنها تمثل ثلاث مجتمعات لها نفس المعالم.

H_1 : أن الفرق بين المسافات في العينات الثلاث هو فرق جوهري، أي أنها تمثل مجتمعات مختلفة.

(٢) مستوى الدلالة $\alpha = 0,05$

(٣) يكون وضع الاختبار في الصورة الاحصائية

$$H = \frac{12}{n(n+1)} \times \text{محي} - \frac{2}{n} \times 3 - (n+1)$$

(٤) حيث أن: عدد العينات = ٣، وعدد المفردات (القياسات) في كل عينة أكثر من ٥ مفردات فإن قيمة «هـ» لها توزيع احتمالي يتطابق مع توزيع مربع كاي بدرجات حرية $n - 1$ حيث n هي عدد العينات. ودرجات الحرية لهذا المثال $2 = 3 - 1$.

(٥) منطقة الرفض: على أساس ما سبق نجد أن قيمة «هـ» الحرجة التي تحدد منطقة الرفض من جدول توزيع كاي هي ٥,٩٩ بدرجات الحرية ٢ ونرفض H_0 فقط عندما تكون قيمة «هـ» المحسوبة من البيانات المشاهدة أكبر من أو تساوي القيمة ٥,٩٩.

(٦) حساب قيمة «هـ» من واقع البيانات المشاهدة: ويتم ذلك عن طريق ترتيب البيانات للعينات الثلاث مجتمعة ترتيباً تصاعدياً أو تنازلياً (من الأقل إلى الأعلى أو من الأعلى إلى الأقل) أى إعطاء كل مفردة من المفردات رتبة خاصة بها، وفي حالة تطابق بعض المفردات يعطى لكل منها رتبة تساوى متوسط رتب هذه المفردات. فمثلاً المفردة ٢٧ تكررت مرتين وأعطيت لها الرتبة ٩,٥ بدلاً من الرتبتين ٩، ١٠ (أى $9 + 10 \div 2 = 9,5$). وبعد إتمام عملية ترتيب المفردات تجمع رتب مفردات كل عينة على حدة، ثم يربع مجموع رتب كل عينة ويقسم على عدد مفرداتها وأخيراً يجمع خارج عمليات القسمة للعينات فتحصل على مجموع متوسطات مربعات الرتب للعينات.

$$H = \frac{12}{19 \times 18} \times \left(\frac{2(87)}{6} + \frac{2(52)}{6} + \frac{2(32)}{6} \right) - \frac{2}{19} \times 3 - (19+1) = 0,35$$

$$= 0,35 - (1261,5 + 450,66 + 170,66) \times 0,35 = 0,35 - 570,66 = -567,31$$

$$٥٧ - (١٨٨٢,٨٢ \times ,٠٣٥) =$$

$$٨,٩٠ =$$

(٧) الاستنتاج: قيمة «هـ» المحسوبة من البيانات المشاهدة أكبر من مثيلتها النظرية (٥,٩٩) بدرجات الحرية ٢ ولذلك نرفض فرض العدم القائل بأنه لا توجد فروق جوهرية بين المسافات (بالكيلو متر) التي قطعها زوار الحدائق الثلاث، ونستنتج أن هناك فروقاً جوهرية بين العينات الثلاث من المسافات بما يدل على أن هذه العينات تمثل إختلافات حقيقة بين المجتمعات التي أتى منها الزوار، وذلك بمستوى دلالة ٠,٠٥.

مثال (٣):

لدراسة خصائص حجم الرواسب الشاطئية وبيان ما إذا كانت طبيعة هذه الرواسب تدل على أنها من أصل مشترك، أو تؤدي إلى التعرف على العامل المرسب وعلى الظروف الجغرافية التي كانت سائدة أثناء الإرساب. وقد اختيرت عشوائياً لذلك ستة من الشواطئ وأخذ من كل منها عينة من الرواسب (في منتصف المسافة تقريباً من المنطقة التي تتعرض لمتوسط المد والجزر). وبعد تجفيف العينات الست وتعرضها لعملية النخل الجاف الميكانيكي تم وزن أحجام الرواسب (بالجرام) وتسجيل نسبتها المختلفة حسب مقياس وحدات فاي (Φ)^(١) ووضعت في الجدول التالي:

(١) مقياس فاي (Φ) هو وسيلة لتصنيف الرواسب حسب حجم حبيباتها. وهو مأخوذ عن نظام وتورث الملليمترى Wentworth's millimetre system الذى يقوم على أساس لوغاريتمى. أى أنه عبارة عن التحويل اللوغاريتمى لحجم الرواسب بالملليمتر، أو هو يساوى اللوغاريتم السالب (للأساس ٢) لحجم الرواسب بالملليمتر (Φ) = - ٢ لو ر، حيث ر = قطر الحبيبات بالملليمتر).

جدول رقم (١٠ - ١٣): وزن حجم الطيات (بالجرام) لست عينات من الرواسب الشاطئية

الشاطئ	أ	ب	جـ	د	هـ	و
وحدات الحجم	الوزن الرتبة	الوزن الرتبة	الوزن الرتبة	الوزن الرتبة	الوزن الرتبة	الوزن الرتبة
٢- إلى ١	١٠٥	٢٨٢	١٦	٣٠٧	٠٠	٢٩١
١- إلى صفر	١٥	٢٥	٠٠	٨٥	٤	٣٢٤
صفر إلى ١	٠	٢٦	٠٠	٢٠١	١٠٥	١١٢
١ إلى ٢	١٤٥	٢٥٢	٥٥	١٨٤	٤٥	١٠٨
٢ إلى ٣	٧٨	١٤٨	١٢١	٣٥٥	٣٥٥	١٦٥
٣ إلى ٤	١٦٠	١٣١	٧٧	١٧٥	٠٠	٢٥
٤ إلى ٥	٨٢	٢٧٣	٦٧	٤	٢,٥	٠٠
الجميع	٧٨٥ = ج ٦ = ن	١٧١ = د ٧ = ن	٤٣,٥ = هـ ٥ = ن	١٤٢,٥ = و ٥ = ن	٥١,٥ = ز ٦ = ن	١٤٣ = ح ٦ = ن

فتكون:

$$\frac{2(43,0)}{0} + \frac{2(171)}{7} + \frac{2(78,0)}{6} \times \frac{12}{19 \times 18} = (هـ)$$

$$19 \times 3 - \left(\frac{2(143)}{6} + \frac{2(51,0)}{0} + \frac{2(142,0)}{6} + \right.$$

$$108 - (12900,769 \times ,0095) =$$

$$14,600 =$$

٧- الاستنتاج: تحت شروط الاختبار نجد أن قيمة «هـ» المحسوبة أكبر من القيمة ١١,٠٧ بمستوى الدلالة ٠,٥، ودرجة الحرية ٥ ولذلك يرفض فرض العدم القائل بأنه لا يوجد فرق جوهري بين حجم الرواسب في العينات الست، أو أن الفرق بينها لا يرجع إلى الصدفة المطلقة بل هو فرق معنوي يؤكد وجود اختلافات حقيقية بين مجتمعات الرواسب الشاطئية التي أخذت منها العينات. ولكن إذا أخذنا مستوى الدلالة ٠,١، كأساس للاختبار فإن قيمة «هـ» الحرجة التي تتحدد منطقة الرفض بنفس درجات الحرية هي ١٥,٠٩، وبالتالي تكون قيمة «هـ» المحسوبة أصغر من هذه القيمة مما يترتب عليه قبول فرض العدم السابق. ويعنى قبول فرض العدم في هذه الحالة أن الفرق بين أحجام الرواسب في العينات يرجع إلى الصدفة المطلقة والناج من خطأ المعاينة. وهذا يعنى بشكل آخر بالنسبة للحالة الأولى بأن هناك احتمالاً مقداره ٩٥% بأن لا يكون الاختلاف بين أحجام الرواسب قد حدث بفعل الصدفة، أما بالنسبة للحالة الثانية، فإن الاختبار يعنى أننا غير متأكدين باحتمال مقداره ٩٩% أن هذا الاختلاف لم يحدث بفعل الصدفة. وفي مثل هذه الحالات يجب على الباحث أن يجمع بيانات أخرى عن الرواسب ولكن من عينة أكبر من العينة السابقة.

وكما لاحظنا في المثالين (٢)، (٣) الأخيرين أن بكل منها عدد من الرتب

المتكافئة (المتساوية) Tied ranks، وفي مثل هذه الحالات لابد من تصحيح قيمة «هـ» المحسوبة حتى لاتتأثر النتائج المترتبة عليها. ويتم تصحيح قيمة «هـ» المحسوبة بقسمتها على عامل التصحيح التالى:

$$\text{عامل تصحيح قيمة هـ} = 1 - \frac{\text{مجموع } (K - 3K)}{n - 3n}$$

حيث ك هي عدد المفردات ذات الرتب المتكافئة بين مفردات كل عينة على حدة، ن هي العدد الكلى للمفردات فى كل العينات. وتطبيق عامل التصحيح السابق على كل من قيمتى «هـ» المحسوبتين فى المثالين (٢)، (٣) نجد أن:

أولاً: فى المثال رقم (٢) يوجد مفردتان فقط متساويتان فى الرتبة وبذلك فإن عامل التصحيح فى هذه الحالة يساوى:

$$0,999 = \frac{1}{5814} - 1 = \frac{[2 - 3(2)]}{18 - 3(18)} - 1$$

وتصبح قيمة «هـ» (٨,٩٠) المصححة:

$$8,91 = \frac{8,90}{0,999} = \text{المصححة «هـ»}$$

ثانياً: فى المثال رقم (٣) يوجد ثلاث عينات بكل منها مفردتان متساويتان فى الرتبة وبذلك فإن عامل التصحيح فى هذه الحالة يساوى:

$$= 1 - \frac{[2 - 3(2)] + (2 - 3(2)) + (2 - 3(2))}{30 - 3(30)}$$

$$,9996 = \frac{18}{42840} - 1 =$$

وتصبح قيمة «هـ» (١٤,٦٠٥) المصححة هي:

$$١٤,٦١١ = \frac{١٤,٦٠٥}{,9996} = \text{المصححة «هـ»}$$

وبلاحظ أن تأثير عامل التصحيح على قيمة «هـ» فى المثالين صغيراً جداً ويكون ذلك صحيحاً حتى إذا وجد عدد كبير من المفردات المتكافئة فى رتبها بين المفردات الكلية للعينات. والفرض الأساسى من عامل التصحيح هو جعل قيمة «هـ» أكبر من القيمة المحسوبة لها مما يزيد من فرصة رفض فرض العدم. أو بعبارة أخرى إذا أهملنا عامل التصحيح السابق، وبصفة خاصة إذا كان ترتيب البيانات ينتج عنه أن ٢٥٪ أو أكثر من المفردات تأخذ رتباً متكافئة، فإن نتيجة اختبار «هـ» تصبح مضللة ويجب معالجتها إحصائياً بكل الحذر والحيلة، أما إذا قل كثيراً عدد الرتب المتساوية بين المفردات فإن تأثير عامل التصحيح يصبح بسيطاً جداً، كما لاحظنا، وبالتالي يمكن إهماله.

الباب الرابع

أساليب قياس العلاقات والتغيرات

مقدمة

الفصل الحادى عشر: تحليل الارتباط

الفصل الثانى عشر: تحليل الإنحدار

الفصل الثالث عشر: تحليل السلاسل الزمنية

مقدمة:

إن مهمة الباحث الجغرافى تنحصر فى معالجة ووصف بيانات المتغيرات (الظواهر) الطبيعية والبشرية الموجودة على سطح الأرض لدراسة مواقعها ومقارنة توزيعاتها والعلاقات التى تربطها البعض. وقد شرحنا فى فصول الأبواب لسابقة من هذا الكتاب الأساليب الكمية الخاصة بوصف وقياس وتحديد درجة الاختلاف بين عينتين أو أكثر على أساس تحديد معالم وخصائص متغير واحد (ظاهرة واحدة) فقط غير أن أسلوب التحليل الكمي لا يقتصر على تحليل كل متغير بمفرده أو فى صورة مستقلة عن المتغيرات الأخرى، بل يعطى لنا نظرية كاملة فى مجال العلاقة بين متغيرين وقياس الترابط بينهما، ولقد كانت وما زالت مقارنة خرائط التوزيعات الجغرافية من أهم الوسائل لدراسة الارتباط والتلازم بين المتغيرات الجغرافية المختلفة. غير أن كثيراً من الدراسات والبحوث الجغرافية تظهر هذه الارتباطات على أنها علاقات سببية، مما كان له أكبر الأثر فى توجيه اللوم للجغرافيين واتهامهم بأنهم كانوا يطوفون بأعينهم على الخرائط ليخرجوا منه باستنتاجات سريعة، أو كانوا يحلقون فى سماء مكونات البيئة ونشاط الإنسان للتدليل على كثير من الحقائق أثبت التحليل الكمي الحديث عدم دقة وخطورة الاعتماد عليها فى وصف ودراسة العلاقات المختلفة والمتشابكة بين المتغيرات الجغرافية المتعددة.

وفى هذا الباب والذى يشتمل على ثلاثة فصول - سنحاول إلقاء الضوء على الأساليب الكمية التى تقيس درجة واتجاه العلاقة بين المتغيرات الجغرافية فى الفصلين الأول والثانى. وتتمثل هذه الأساليب فى تحليل الارتباط وتحليل الانداز. أما فى الفصل الأخير فستعرض لدراسة بيانات الظاهرة الجغرافية على مدى فترات زمنية متتابة (أو ما يعرف بتحليل السلاسل الزمنية) بغرض وصف خط سير (أى تطور أو نمو) هذه الظاهرة فى مدى زمنى محدد. ثم نتعرف على الأساليب المتبعة فى قياس التغيرات المختلفة التى قد تطرأ على تطور الظاهرة بفعل المؤثرات العديدة والمتباينة بهدف الاستفادة من هذه الأساليب فى التنبؤ بما يمكن أن تكون عليه بيانات الظاهرة قيد التحليل سواء فى الماضى أو فى المستقبل.

الفصل الحادي عشر تحليل الارتباط

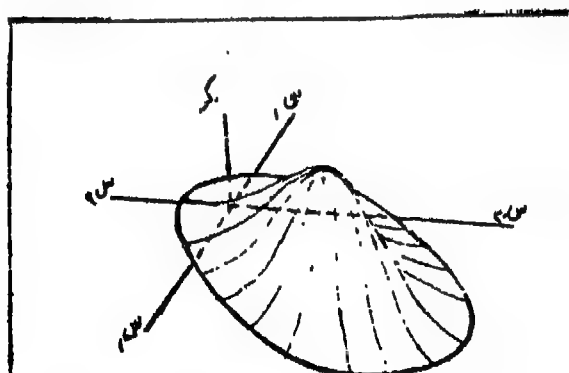
الفصل الحادى عشر

تحليل الارتباط

سبق القول بأن الأسلوب الكمي الحديث يساعد الباحث الجغرافى على الوصول إلى أهدافه العلمية بوسائل أكثر دقة من الأسلوب الوصفى التقليدى. ومن هنا نجد أن الجغرافيين الآن يهتمون بتطبيق أسلوباً كمياً معيناً - هو تحليل الارتباط - على بيانات الظواهر (المتغيرات) الجغرافية لكي يعرفوا به إن كان ثمة علاقة أو ارتباط بين ظاهرتين معينتين ولتحديد ما إذا كانت هذه العلاقة تعود إلى تلازم بين الظاهرتين أو إلى اختلافات ترجع إلى الصدفة المطلقة نتيجة خطأ المعاينة. ونقصد بتحليل الارتباط أنه الأسلوب الذى يقيس درجة الترابط بين ظاهرتين (متغيرين) إذا كانت العلاقة بينهما ليست علاقة دالية (أى أن التغير فى أحد الظاهرتين بالتلازم - لعلاقة بين طول النهر وعرضه فإن أى تغير فى أحدهما لا يسبب تغيراً فى الآخر ولكن المتغيرين مرتبطين بحجم الحوض النهري وخصائصه الجيومورفولوجية).

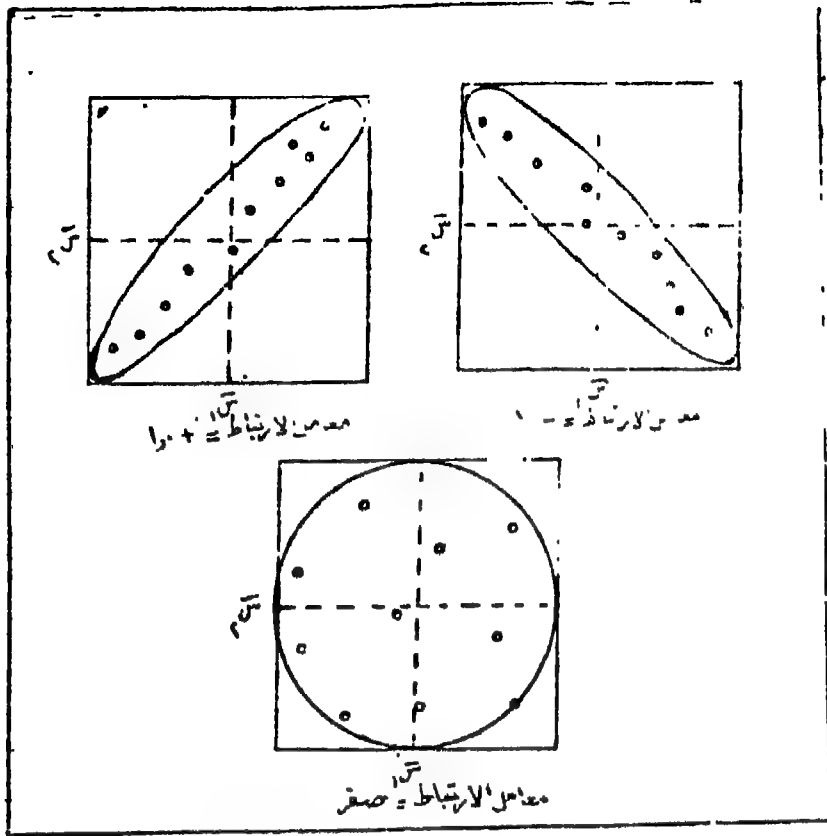
ويشترط عند تحليل الارتباط أن يتبع توزيع كل متغير من المتغيرين التوزيع المعتدل، أما توزيع المتغيرين معاً فإنه يشترط أن يتبع توزيع Bivariate Normal Distribution (شكل رقم ١١-١)، الذى يشبه الناقوس حيث تمثل قاعدته المتغيرين وارتفاعه يمثل التكرار أو لكلا المتغيرين، وذلك حتى يمكن تطبيق الأساليب البارامترية الخاصة بقياس درجة الارتباط. بين المتغيرين، بينما إذا كان توزيع أحد المتغيرين، أو كلاهما، لا يتبع التوزيع المعتدل فإنه يمكن تطبيق أساليب أخرى غير بارامترية لقياس درجة الارتباط بين المتغيرين. ويلاحظ من الشكل رقم (١١-١) أن أى قطع عمودى على المستوى الذى يمثل المتغير الأول (س_١) ينتج عنه منحنى معتدل يمثل توزيعاً للمتغير الثانى (س_٢) والعكس صحيح. أما إذا كان القطع موازياً للمستويين س_١، س_٢ فإن ذلك يعطى شكل قطع ناقص Ellips يمكن اعتباره دليلاً على نوع وقوة الإرتباط (العلاقة) بين المتغيرين (شكل رقم ١١-١). فإذا

كان القطع الناتج يأخذ اتجاهاً معيناً فإن ذلك يدل على نوع الارتباط فيما يكون الارتباط موجب (أى علاقة طردية)؛ أى أن تزايد قيم أحد المتغيرين يصحبه تزايد فى قيم المتغير الآخر والعكس - وهناك متغيرات كثيرة تتبع هذا النوع من الارتباط نذكر منها على سبيل المثال لا الحصر تزايد الأمطار والإنتاج الزراعى فى المناطق الجافة وشبه الجافة؛ بمعنى أن أى تزايد فى الأمطار يصحبه تزايد فى إنتاج المحاصيل، شدة انحدار السفوح والتعرية؛ إذا أنه كلما زادت درجة الانحدار زادت شدة التعرية، زيادة سرعة المياه فى الأنهار وكمية الرواسب المحمولة، والكفاءة الإنتاجية للعمال والإنتاج الصناعى؛ فكلما تحسنت الكفاءة الإنتاجية كلما زاد الإنتاج وكلما ضعفت الكفاءة قل الإنتاج. وإما يكون الارتباط سالب (علاقة عكسية)؛ أى تزايد قيم أحد المتغيرين يصحبه انخفاض فى قيم المتغير الآخر. ومن أمثلة هذا النوع من العلاقة العكسية كثافة السكان والبعد عن وسط المدينة، كذلك أسعار الأراضى والبعد عن قلب المدينة التجارى - فالواضح أن تزايد المسافة بالبعد عن وسط المدينة وقلبها التجارى يرتبط به تغير عكسى فى كثافة السكان وأسعار الأراضى، أى أن الكثافة وأسعار الأراضى تقل كلما تزايد طول المسافة عن وسط المدينة وإما يكون الارتباط معدوم؛ أى أنه ليس هناك علاقة أو ارتباط (موجب أو سالب) بين المتغيرين ون أمثلة ذلك البعد عن وسط المدينة وكمية الإنتاج الصناعى، أو البعد عن قلب المدينة التجارى وسرعة المياه فى المجارى النهرية.



شكل رقم (١١-١): التوزيع المعتدل للمتغيرين

١١، ١٢ معاً Bivariate Normal Distribution



شكل رقم (١١-٢): قوة الارتباط بين المتغيرين S_1 ، S_2
كما يوضحها شكل إنتشار المفردات لكل منهما

وكما يدل اتجاه القطع الناقص - الذى ينتج عن قطع توزيع المتغيرين معا قطعاً موازياً للسطحين S_1 ، S_2 فى الشكل رقم (١١-١) على نوع الارتباط أو العلاقة، يدل شكل القطع الناقص نفسه على قوة الارتباط أو العلاقة. فإذا كان شكل القطع الناقص ضيق ومحدود، أى أن بيانات المتغيرين تقع فى مجال انتشار متقارب أو على امتداد خط مستقيم، فهذا يدل على أن هناك علاقة قوية بين المتغيرين. أما إذا كان شكل القطع بعض الشيء بحيث تظهر البيانات متباعدة تباعداً

طفيفاً ولكن حول خط مستقيم، دل ذلك على وجود علاقة ضعيفة. أما في حالة إذا كان شكل القطع دائري، أى أن هناك تباعداً كبيراً بين البيانات بحيث يتعذر وقوعها على امتداد خط مستقيم، فإن ذلك يدل على عدم وجود علاقة بين المتغيرين أو يدل على استقلال كل منهما عن الآخر.

وما تجدر الإشارة إليه بشأن العلاقات الارتباطية واتجاهها بين المتغيرات، أنه في حالة إثبات وجود علاقة قوية أياً كان نوعها بين متغيرين، أو بعبارة أخرى أنه لو تعرفنا خلال الاختبار بأن العلاقة بين متغيرين لانهود إلى الصدفة فإن هذا لايعنى بأن هناك علاقة سببية بينهما، ذلك لأن وجود الارتباط لايعنى بالضرورة وجود علاقة سببية (علاقة تبعية مباشرة) بين المتغيرين، بل أن كل ما يكشف عنه (الإختبار) هو أن المتغيرين متلازمان تلازماً شديداً، مما يتيح الفرصة ويفسح المجال بعد ذلك البحث عن العلاقة الحقيقية الواقعة بينهما، ولكن إذا كان هناك علاقة سببية بين المتغيرين فلا بد أن يكون هناك ترابط بينهما. فمثلاً عند حساب الارتباط بين متوسط محصول القمح ودرجة الإصابة بدودة ورق القطن وجد أن هناك ارتباطاً قوياً بينهما، ولكن من الصعب تفسير هذه العلاقة من الناحية النطقية إذ لا يوجد سبب واحد بين هذين المتغيرين يؤدي إلى وجود هذه العلاقة، ولكن هذا الارتباط (العلاقة) الزائف سببه الظروف المناخية الملائمة للنمو نبات القمح أثناء فصل الشتاء تؤثر على أعداد دودة ورق القطن. كذلك قد يكون هناك ارتباط قوى بين عدد الكتب المنشورة في كل سنة وعدد مباريات كرة القدم في كل سنة، مثل هذا الارتباط يشار إليه بأنه ارتباط لاعمى له أو ارتباط زائف. ومن هنا يمكن القول أن الترابط ليس شرطاً للعلاقة السببية ولكن السببية شرط للترابط.

مقاييس الارتباط Measures Of Correlation

يمكن أن نحدد بصورة وصفية مدى جودة وصف العلاقة أو الارتباط بين المتغيرات بملاحظة الشكل البياني الذى يوضح مجال انتشار قيم المتغيرين فى نظام الاحداثيات المتعامدة، وهو الشكل الذى يعرف باسم «شكل الانتشار» (راجع لفصل الثالث الخاص بطرق العرض البياني للبيانات الاحصائية). ولكن يمكن معرفة مدى جودة تعبير خط مستقيم عن العلاقة بين المتغيرات عن طريق حساب

معامل ارتباط Correlation Coefficient والذي وعلى أساسه يستخلص الباحث الجغرافى النتائج ويتخذ القرارات الخاصة بتوضيح العلاقات بين المتغيرات (الطبيعية أو البشرية) الجغرافية.

حساب معامل الارتباط:

يستخدم مصطلح «معامل الارتباط» ليعنى الارتباط الخطى (أو العلاقة الخطية) بين متغيرين، وهو لذلك يستخدم - كمقياس إحصائى - لتحديد نوع العلاقة وقوتها بين المتغيرات. وتتراوح قيمة معامل الارتباط المحسوبة بين $(-1, 0)$ ، $(0, +1)$ ، حيث تشير القيمة $(-1, 0)$ إلى وجود ارتباط سالب أو عكسى تام Negative (Inverse) Correlation، أما $(+1)$ فترمز إلى وجود علاقة ارتباط طردية أو موجبة تامة. Positive (Direct) Correlation والقيمتان تدلان على أن جميع القيم الممثلة للعلاقة بين المتغيرين تقع على خط مستقيم. وكلما أخذت القيم تنحرف عن الخط المستقيم كلما قلت قيمة معامل الارتباط عن القيمتين السابقتين بحكم ضعف العلاقة بين قيم المتغيرين، حتى إذا وصلت إلى درجة الصفر دل ذلك على انعدام العلاقة الارتباطية. وإذا كانت الإشارات $(+)$ ، $(-)$ تشير إلى نوع الارتباط الخطى، فإن قيمة معامل الارتباط لا تميز لها أى أنها لا تعتمد على وحدات القياس المستخدمة. وينبغى أن نؤكد هنا، مرة أخرى على أن معامل الارتباط يقيس مدى جودة توفيق الصيغة الاحصائية المفترضة للبيانات، أو بعبارة أخرى أن قيمة معامل الارتباط المحسوبة بين متغيرين تقيس فقط درجة العلاقة بينهما بالنسبة إلى نوع الصيغة الاحصائية المستخدمة.

وهناك صيغ مختلفة لحساب قيمة معامل الارتباط (يرمز له بالرمز «ر») بين متغيرين، إلا أن الصيغة التالية يفضلها كثير من الاحصائيين لأنها تعتمد على القيم الأصلية للمتغيرين (س_١، س_٢).

$$r = \frac{\sum (س_١ - \text{مجد } س_١) (س_٢ - \text{مجد } س_٢)}{\sqrt{[\sum (س_١ - \text{مجد } س_١)^2] \times [\sum (س_٢ - \text{مجد } س_٢)^2]}}$$

حيث ن هي عدد أزواج القيم للمتغيرين معاً.
وسنوضح كيفية حساب معامل الارتباط باستخدام هذه الصيغة من المثال
التالى:

مثال (١):

لدراسة العلاقة بين شكل الشاطئ ممثلاً بدرجات الانحدار، وشدة التعرية
البحرية ممثلة بقوة الأمواج، أخذت عينة مكونة من ٦ قطاعات على الشاطئ
وقيست درجة الانحدار وقوة الأمواج فكانت كمايلي:

٢٨	٢٤	٢٠	١٨	١٥	١٠	قوة الأمواج (عشرات كجم/م
٦	٥	٤	٤	٣	٢	انحدار الشاطئ (بالدرجات)

والمطلوب حساب مقدار الارتباط بين هاتين الظاهرتين:
لتسهيل عملية الحساب يتم ترتيب البيانات فى الجدول التالى:

جدول رقم (١١-١)

حساب الارتباط بين قوة الأمواج وانحدار الشاطئ

قوة الأمواج	درجة الإنحدار		س ١	س ٢	س ٣
	س ١	س ٢			
١٠	٢	١٠٠	٤	٢٠	٥٠٥
١٥	٣	٢٢٥	٩	٤٥	٥٧٦
١٨	٤	٣٣٤	١٦	٧٢	٧٨٤
٢٠	٤	٤٠٠	١٦	٨٠	٧٨٤
٢٤	٥	٥٧٦	٢٥	١٢٠	٧٨٤
٢٨	٦	٧٨٤	٢٦	١٦٨	٧٨٤
١١٥	٢٤	٢٤٠٩	١٠٦	٥٠٥	٧٨٤

معامل الارتباط (ر) =

$$\frac{N \text{ مج-} 1 \text{ س} 1 \text{ س} 2 - (\text{مج س} 1) (\text{مج س} 2)}{\sqrt{[N \text{ مج-} 1 \text{ س} 1 \text{ س} 2 - (\text{مج س} 1) (\text{مج س} 2)]^2}}$$

وحيث $N = 6$ فإن:

$$r = \frac{24 \times 115 - 500 \times 6}{\sqrt{[2(24) - (106 \times 6) \times 2(115) - 2409 \times 6]}}$$

$$r = \frac{2760 - 3020}{\sqrt{(576 - 636) (13225 - 14454)}}$$

$$r = \frac{270}{\sqrt{73740}} = \frac{270}{60 \times 1229}$$

$$r = \frac{270}{271,6} = 0,9941 +$$

أى أن معامل الارتباط هو $+ 0,9941$ قريب جداً من قيمة الارتباط التام $(+1)$ مما يدل على أن معامل الارتباط قوى جداً. ويمكن أن نستنتج من ذلك أن هناك علاقة جغرافية وثيقة أو شديدة بين عنصرى قوة الأمواج وطبيعة شكل الشاطئ. ويلاحظ هنا أن العلاقة موجبة (علاقة طردية) أى أنه عندما تزداد قوة الأمواج يشتد انحدار الشاطئ.

مثال (٢):

فى دراسة لبيان العلاقة بين كثافة السكان والبعد عن قلب المدينة التجارى، قام باحث جغرافى بقياس المسافة من قلب المدينة (بالكيلومتر) وحساب كثافة السكان فى الكيلومتر المربع من قلب المدينة حتى خارجها فكانت البيانات التى حصل عليها كمايلى:

المسافة من قلب المدينة (كم)	١	٢	٣	٤	٥	٦	٧	٨
الكثافة (ألف شخص / كم ^٢)	١٠	٦	٨	٦	٥	٣	٣	٢

ولحساب معامل الارتباط بين هذين المتغيرين ترتب البيانات كما فى الجدول التالى:

جدول رقم (١١-٢)

حساب العلاقة بين كثافة السكان والمسافة من قلب المدينة

المسافة	الكثافة		٢ ^٢ مس	١ ^٢ مس
	١ مس	٢ مس		
١	١٠	١	١٠٠	١٠
٢	٦	٤	٣٦	١٢
٣	٨	٩	٦٤	٢٤
٤	٦	١٦	٣٦	٢٤
٥	٥	٢٥	٢٥	٢٥
٦	٢	٣٦	٩	١٨
٧	٣	٤٩	٩	٢١
٨	٢	٦٤	٤	١٦
المجموع ٣٦	٤٣	٢٠٤	٢٨٣	١٥٠

وبتطبيق معادلة معامل الارتباط السابقة (حيث ن = ٨) نحصل على قيمة المعامل وهى:

$$r = \frac{43 \times 36 - 150 \times 8}{\sqrt{[2(43) - (283 \times 8) \times 2(36) - 204 \times 8]}}$$

$$\frac{348 -}{139440} \sqrt{\quad} = \frac{348 -}{410 \times 336} \sqrt{\quad} =$$

$$0,9319 - = \frac{248 -}{373,42} \sqrt{\quad} =$$

يتضح من نتيجة المعادلة أن هناك علاقة ارتباط قوية بين المسافة من قلب المدينة وكثافة السكان، ولكنها علاقة عكسية (سالبة)، أى أنه عندما تزداد المسافة من قلب المدينة تقل كثافة السكان فى الكيلو متر المربع تبعاً لذلك.

معامل ارتباط ضرب العزوم Product Moment Correlation Coefficient

يعد معامل ارتباط ضرب العزوم أو ما يعرف باسم معامل ارتباط بيرسون - Pearson Correlation Coefficient من أقوى الأساليب الإحصائية البارامترية (المعلمية) التى تقيس العلاقة بين متغيرين (ظاهرتين) يشترط فى بياناتها أن تكون من بيانات الفترة - Scaled - Intervally.

وعلى الرغم من أن صيغة بيرسون لحساب معامل الارتباط تعتبر كشفاً علمياً له أهمية كبيرة - لتحديد نوع ودرجة العلاقة بين المتغيرات - فى ميدان العلوم الطبيعية والبشرية، إلا أنها تحتاج إلى عمليات حسابية متعددة ودقيقة مما قد يؤدى إلى ارتكاب بعض الأخطاء التى قد تؤثر كثيراً على النتائج النهائية للدراسة. ولكن بفضل استخدام أجهزة الحاسبات الآلية Computers فقد أصبحت العمليات الحسابية لهذه الصيغة تتم بسهولة ويسر وبدون الوقوع فى أخطاء، مما أدى إلى أنها أصبحت تستخدم بكثرة فى البحوث الجغرافية.

وتعتمد أساساً صيغة بيرسون لمعامل الارتباط للعينة على حساب انحرافات قيم المتغيرات عن متوسطاتها الحسابية، وتكتب الصيغة بالشكل التالى:

$$r = \frac{\frac{1}{n} [\sum (x_1 - \bar{x}_1) \times (x_2 - \bar{x}_2)]}{\sqrt{\frac{\sum (x_1 - \bar{x}_1)^2}{n} \times \frac{\sum (x_2 - \bar{x}_2)^2}{n}}}$$

وحساب معامل الارتباط بهذه الصيغة يكون صعباً، خاصة إذا كانت قيمة المتوسط الحسابي تحتوى على كسور عشرية مما قد يؤدي إلى تعقيد العمليات الحسابية وبالتالي يزيد من احتمالات الخطأ في النتيجة النهائية، لذلك فقد اشتقت عدة صيغ أخرى تكتب على النحو التالي:

$$r_{pb} = \frac{\text{مجم } ١س - ٢س \quad \text{ن } ١س - ٢س}{\sqrt{(\text{مجم } ١س - ٢س)^2 (\text{ن } ١س - ٢س)^2}} \dots\dots (٣-١١)$$

أو:

$$r_{pb} = \frac{\text{مجم } ١س - ٢س}{\text{ن}} \times \frac{\sqrt{\frac{\text{مجم } (١س)^2}{\text{ن}} - \frac{\text{مجم } (٢س)^2}{\text{ن}}}}{\sqrt{\frac{\text{مجم } (١س)^2}{\text{ن}} - \frac{\text{مجم } (٢س)^2}{\text{ن}}}} \dots\dots (٤-١١)$$

ويلاحظ أن حساب معامل الارتباط من المعادلات السابقة يتطلب: حساب المتوسط الحسابي والانحراف المعياري للمتغيرين (١س ، ٢س)، ومجموع حاصل ضرب كل من المتغيرين (١س ، ٢س).

وهناك صيغة أخرى لحساب معامل الارتباط تعرف بالطريقة المختصرة التي تستفيد من الخصائص الحسابية لمعامل الارتباط والتي يمكن أن تقلل إلى حد كبير من العمليات الحسابية والوقوع في الخطأ، هذه الخصائص هي: أن قيمة معامل الارتباط لا تتغير إذا قسمنا أو ضربنا جميع قيم المتغير الأول على أو في عدد ثابت، أو إذا قسمنا أو ضربنا جميع قيم المتغير الثاني على أو في أى عدد ثابت آخر. كما أن قيمة معامل الارتباط لا تتغير إذا أضفنا أو طرحنا أى عدد ثابت إلى أو من جميع قيم المتغير الأول، أو إذا أضفنا أو طرحنا أى عدد ثابت آخر إلى أو من جميع قيم المتغير الثاني. وتكتب صيغة بيرسون المختصرة لتسهيل حساب معامل الارتباط على النحو التالي:

$$\text{بب} = \frac{\frac{1}{n} \cdot (\text{مجم} - \bar{X}_1 \times \bar{X}_2 - \bar{X}_1 \times \bar{X}_2)}{(11-5) \dots} \times \sqrt{\frac{\text{مجم}^2 - \bar{X}_1^2}{n}} \times \sqrt{\frac{\text{مجم}^2 - \bar{X}_2^2}{n}}$$

حيث \bar{X} هي انحراف قيم كل من المتغيرين عن المتوسط الحسابي لكل منهما، \bar{X} هي متوسط الانحراف لقيم كل من المتغيرين عن المتوسط الحسابي لكل منهما، n هي عدد أزواج المفردات للمتغيرين معاً.

ولحساب قيمة معامل الارتباط (r) بواسطة الصيغ السابقة نأخذ الأمثلة الآتية:

مثال (٢):

البيانات التالية تمثل مساحة حوض النهر وطوله لعينة من عشرة أنهار لأحد نظم التصريف النهري في منطقة ما. والمطلوب حساب معامل الارتباط بين هذين المتغيرين.

طول النهر (كم) ٩،٤٠، ١٠،٨٥، ١٠،٦٩، ١٠،٩٧، ١١،٩٤، ١٨،٦٩، ١٧،٢٣، ١٠،٦٨، ١٠،١٤، ١٨،٤١.

مساحة حوض النهر (كم) ٤،٠٨، ٤،٥٣، ٤،٩٢، ٦،١٨، ٦،٢٢، ٦،٥٤، ٧،٥٥، ١٢،٥٠، ١٠،١٦، ٨،٩٢.

ولتوضيح خطوات حساب معامل ارتباط بيرسون بالصيغة (١١-٢) يتم ترتيب المعلومات كما في الجدول التالي:

جدول رقم (١١-٣)

طريقة حساب معامل الارتباط بين طول النهر (س_١) ومساحة حوضه (س_٢)

$\frac{\times (س_1 - \bar{س_1})}{(س_2 - \bar{س_2})}$	$\bar{س_2} - س_2$	$\bar{س_1} - س_1$	$\bar{س_2} - س_2$	$\bar{س_1} - س_1$	$س_2$	$س_1$
١٦,٩٤	٩,٤٩	٣٠,٢٥	٣,٠٨	٥,٥٠	٤,٠٨	٩,٤٠
١٠,٦٥	٦,٩٢	١٦,٤٠	٢,٦٣	٤,٠٥	٤,٥٣	١٠,٨٥
٩,٤٣	٥,٠٢	١٧,٧٢	٢,٢٤	٤,٢١	٤,٩٢	١٠,٦٩
٣,٨٥	٠,٩٦	١٥,٤٤	٠,٩٨	٣,٩٣	٦,١٨	١٠,٩٧
٢,٧٢	٣,٦٩	٨,٧٦	٠,٩٢	٢,٩٦	٦,٢٢	١١,٩٤
٢,٣٥	٠,٣٨	١٤,٣٦	٠,٦٢	٣,٧٩	٦,٥٤	١٨,٦٩
٠,٩٠	٠,١٥	٥,٣٤	٠,٧٩	٢,٣٣	٧,٥٥	١٧,٢٣
٥,٧٠	٣,١٠	١٠,٥٠	١,٧٦	٣,٢٤	٨,٩٢	١٨,٤١
١٥,٧٢	٩,٠٠	٢٧,٤٦	٣,٠٠	٥,٢٤	١٠,١٦	٢٠,١٤
٣٠,٨٧	٢٨,٥٢	٣٣,٤١	٥,٣٤	٥,٧٨	١٢,٥٠	٢٠,٦٨
٩٩,١٣	٦٧,٢٣				٧٠,٦٠	١٤٩,٠٠

$$٧,١٦ = \frac{١٤٩}{١٠} = \bar{س_2} , ١٤,٩ = \frac{١٤٩}{١٠} = \bar{س_1}$$

ومن البيانات السابقة يمكن حساب معامل ارتباط العينة (ر) بالمعادلة رقم (١١-٢) كمايلي:

$$\frac{٩٩,١٣}{١٠} \times \frac{١}{\sqrt{\frac{٦٧,٢٣}{١٠} \times \frac{١٧٩,٦٤}{١٠}}} =$$

$$0,902 + = \frac{9,913}{10,989} = \frac{9,913}{120,772\sqrt{}} =$$

وحيث أن قيمة معامل الارتباط (+0,902) قريبة من العلاقة الارتباطية التامة والتي توضح أن العلاقة بين طول النهر ومساحة حوضه هي علاقة موجبة طردية، بمعنى أن هذين المتغيرين يزدادان أو يتناقصان في قيمتهما معاً.

مثال (٣):

البيانات التالية توضح إنتاج الفدان من محصول القمح (بالأردب) في منطقتين زراعتين س١، س٢ خلال عشرة سنوات متتالية هي:

السنة	١٩٧١	٧٢	٧٣	٧٤	٧٥	٧٦	٧٧	٧٨	٧٩	١٩٨٠	المتوسط
المنطقة س١	٢٥	٢٦	٣٤	٢٥	٢٤	٢٨	٢٧	٢٩	٢٨	٢٩	٢,٧٥
المنطقة س٢	٢١	١٧	٣٥	٢١	١٩	٢٢	٢٦	٢٢	٢٦	٢٦	٢٣,٥

يلاحظ على هذه البيانات أن إنتاج الفدان في المنطقة س١ يزيد في السنوات الخمس الأخيرة عن متوسط الإنتاج في الفترة كلها، بينما في المنطقة س٢ يزيد إنتاج الفدان في أربع سنوات فقط عن متوسط الإنتاج في فترة العشرة سنوات. ويمكن إيجاد قيمة معامل الارتباط بين الإنتاج في المنطقتين باستخدام الصيغة رقم (١١-٣) وذلك على النحو التالي:

جدول رقم (١١-٤)

خطوات حساب معامل الارتباط بين إنتاج القمح في المنطقتين س١، س٢

السنة	س١	س٢	س١	س٢	س١ س٢
١٩٧١	٢٥	٢١	٦٢٥	٤٤١	٥٢٥
١٩٧٢	٢٦	١٧	٦٧٦	٢٨٩	٤٤٢
١٩٧٣	٣٤	٣٥	١١٥٦	١٢٢٥	١١٩٠
١٩٧٤	٢٥	٢١	٦٢٥	٤٤١	٥٢٥
١٩٧٥	٢٤	١٩	٥٧٦	٣٦١	٤٥٦
١٩٧٦	٢٨	٢٢	٧٨٤	٤٨٤	٦١٦
١٩٧٧	٢٧	٢٦	٧٢٩	٦٧٦	٧٠٢
١٩٧٨	٢٩	٢٢	٨٤١	٤٨٤	٦٣٨
١٩٧٩	٢٨	٢٦	٧٨٤	٦٧٦	٧٢٨
١٩٨٠	٢٩	٢٦	٨٤١	٦٧٦	٧٥٤
			٧٦٣٧	٥٧٥٣	٦٥٧٦

$$\bar{س١} = ٢٧,٥ ، \bar{س٢} = ٢٣,٥$$

وتكون قيمة معامل الارتباط (ر) للبيانات السابقة باستخدام المعادلة (١١-٣)

هو:

$$r = \frac{٢٣,٥ \times ٢٧,٥ \times ١٠ - ٦٥٧٦}{\sqrt{[٢(٢٣,٥) \times ١٠ - ٥٧٥٣] [٢(٢٧,٥) \times ١٠ - ٧٦٣٧]}}$$

$$= \frac{١١٣,٥}{\sqrt{٢٢٠,٥ \times ٧٤,٥}}$$

$$= \frac{١١٣,٥٠}{١٣١,٠٤٣} = ٠,٨٦٦$$

وتدل قيمة معامل الارتباط المحسوبة (+ ٠,٨٦٦) على وجود علاقة طردية قوية بين إنتاج القمح فى المنطقتين خلال فترة العشرة سنوات قيد الدراسة وتجدد الإشارة هنا إلى أن قيمة معامل الارتباط الذى حصلنا عليها لا توضح السبب فى وجود هذه العلاقة وسوف نتناول بعد قليل كيفية توضيح العلاقات الإحصائية المعنوية بين المتغيرات.

مثال (٤):

فى أحد مراكز البحوث الزراعية التى تهتم بتأثير ارتفاع سطح الأرض عن مستوى سطح البحر على الإنتاج الزراعى سجلت البيانات التالية عن إنتاج الفدان من القمح (بالأردب) وارتفاع الأرض المنتجة (بالأقدام):

الارتفاع بالقدم	١٠٠	٢٠٠	٥٠٠	٧٠٠	٨٠٠	١٠٠٠	١٤٠٠	١٥٠٠	١٨٠٠	٢٠٠٠
-----------------	-----	-----	-----	-----	-----	------	------	------	------	------

(س١)

إنتاج الفدان بالأردب	٣٠	٣٠	٣١	٢٤	٢٦	٢٣	١٣	١٧	١٤	١٢
----------------------	----	----	----	----	----	----	----	----	----	----

(س٢)

والمطلوب حساب معامل الارتباط بين هذين المتغيرين. وإجراء ذلك ترتب البيانات كما فى الجدول التالى:

جدول رقم (١١-٥)
حساب معامل الارتباط بين ارتفاع سطح لأرض وإنتاج القمح

الارتفاع بالقدم (س)	إنتاج القمح بالأرداب س	س _١	س _٢	س _١ س _٢
١٠٠	٣٠	١٠٠٠٠	٩٠٠	٩٠٠٠٠
٢٠٠	٣٠	١٠٠٠٠	٩٠٠	٩٠٠٠٠
٥٠٠	٣١	٢٥٠٠٠٠	٩٦١	٢٥٠٠٠٠
٧٠٠	٢٤	٤٩٠٠٠٠	٥٧٦	١٦٨٠٠٠
٨٠٠	٢٦	٦٤٠٠٠٠	٦٧٦	٢٠٨٠٠٠
١٠٠٠	٢٣	١٠٠٠٠٠٠	٥٢٩	٢٣٠٠٠٠
١٤٠٠	١٣	١٩٦٠٠٠٠	١٦٩	١٨٢٠٠٠
١٥٠٠	١٧	٢٢٥٠٠٠٠	٢٨٩	٢٥٥٠٠٠
١٨٠٠	١٤	٣٢٤٠٠٠٠	١٩٦	٢٥٢٠٠٠
٢٠٠٠	١٢	٤٠٠٠٠٠٠	١٤٤	٢٤٠٠٠٠
المجموع ١٠٠٠٠	٢٢٠	١٣٨٨٠٠٠٠	٥٣٤٠	١٧٨٠٠٠

$$س_٢ = \frac{٢٢٠}{١٠} = ٢٢ ، س_١ = \frac{١٠٠٠٠}{١٠} = ١٠٠٠$$

وباستخدام المعادلة رقم (١١-٤) يمكن الحصول على قيمة معامل الارتباط كمايلي:

$$r = \frac{٢٢ \times ١٠٠٠ - \frac{١٧٨٠٠٠}{١٠}}{\sqrt{٢(٢٢) - \frac{٥٣٤٠}{١٠}} \times \sqrt{٢(١٠٠٠) - \frac{١٣٨٨٠٠٠}{١٠}}}$$

$$\frac{220000 - 17800}{\sqrt{484 - 534} \times \sqrt{1000000 - 1388000}} = \therefore$$

$$\frac{74,04200 -}{\sqrt{7007 \times 722,89}} =$$

$$0,954 - = \frac{4200 -}{4403,83} =$$

وتدل قيمة معامل الارتباط (-0,954) على أن هناك علاقة عكسية قوية بين ارتفاع سطح الأرض عند مستوى سطح البحر وإنتاج القدان من محصول القمح.

مثال (٥):

البيانات التالية توضح كمية الأمطار الساقطة (بالبوصات) وشدة سريان المياه (بالبوصات) في أحد الأنهار بإحدى المناطق ذات المناخ الموسمي لفترة ست عشرة سنة :

كمية الأمطار (بالبوصات): ٤٦,٤، ٦٣,٠، ٤٨,٨، ٦٠,١، ٥٠,٦، ٥٧,٥، ٥٥,٥، ٥٧,٠، ٦٠,٨، ٤٨,٣، ٥٩,٠، ٤١,٠، ٦٦,٧، ٥٦,٤، ٥٨,٣، ٥٥,٧.

شدة سريان مياه النهر (بالبوصات): ٤٦,٨، ٣١,٩، ٤٦,٢، ٣٤,٢، ٤٧,٥، ٣٥,٢، ٤٠,٥، ٤١,٣، ٤٤,٨، ٣٨,٥، ٣٩,١، ٢٩,٥، ٤٦,٥، ٤٣,٤، ٤٠,٩، ٤١,٣.

ولإجراء حساب معامل الارتباط بين هذين المتغيرين، نختار وسطاً فرضياً لكمية المطر (س) وليكن ٥٥ بوصة، ووسطاً فرضياً لشدة سريان مياه النهر وليكن ٤٠ بوصة. وفي مثل هذه الحالة نستخدم الطريقة المختصرة ولتسهيل العمليات الحسابية نتخلص من الكسور العشرية وذلك بالضرب في ١٠ وذلك على النحو التالي:

جدول رقم (١١-٦)

خطوات حساب معامل الارتباط بين كمية الأمطار وشدة سريان مياه النهر

ح ^٢ من _١	ح ^٢ من _١	ح ^٢ من _١ × ح ^٢ من _١	ح ^٢ من _٢ (١٠ - ٢ من _٢ - ٤٠)	ح ^٢ من _١ (٥٥ - ١ من _١ - ٥٥)
٦٥٦١	٧٣٩٦	٦٩٦٦	٨١ -	٨٦ -
٤٦٢٤	٦٤٠٠	٥٤٤٠	٦٨ +	٨٠ +
٣٣٦٤	٢٨٤٤	٣٥٩٦	٥٨ -	٦٢ -
٥٦٢٥	٢٦٠١	٣٨٢٥	٧٥ +	٥١ +
٢٣٠٤	١٩٣٦	٢١١٢	٤٨ -	٤٤ -
٢٥	٦٢٥	١٢٥	٥ +	٢٥ +
١٦٩	٢٥	٦٥	١٣ +	٥ +
١٢٢٥	٤٠٠	٧٠٠	٣٥ +	٢٠ +
٢٣٠٤	٢٣٦٤	٢٧٨٤	٤٨ +	٥٨ +
٢٢٥	٤٤٨٩	١٠٠٥	١٥ -	٦٧ -
٨١	١٦٠٠	٣٦٠	٩ -	٤٠ +
١٨٢٢٥	١٩٦٠٠	١٨٩٠٠	١٣٥ -	١٤٠ -
٤٢٢٥	١٣٦٨٩	٧٦٠٥	٦٥ -	١١٧ +
١١٥٦	١٩٦	٤٧٦	٣٤ +	١٤ +
٨١	١٠٨٩	٢٩٧	٩ +	٣٣ +
١٦٩	٤٩	٩١	١٣ +	٧ +
٥٠٣٦٣	٦٧٣٠٢	٥٣٦٢٧	١٩ +	٥١ +
مجموع ح ^٢ من _٢	مجموع ح ^٢ من _١	مجموع ح ^٢ من _١ × ح ^٢ من _١	مجموع ح ^٢ من _٢	مجموع ح ^٢ من _١

وتطبيق المعادلة رقم (١١-٥) يمكن حساب معامل الارتباط (ر) بعد حساب حدود هذه المعادلة كمايلي:

$$\bar{X}_1 = \frac{51}{16} = \frac{\text{مجم ح س}_1}{n} = \bar{X}_1$$

$$\bar{X}_2 = \frac{19}{16} = \frac{\text{مجم ح س}_2}{n} = \bar{X}_2$$

$$\bar{X}_1 \bar{X}_2 = 1.19 \times 3.19 = 3.79$$

$$\bar{X}_1 \bar{X}_2 = \frac{53627}{16} = \frac{\text{مجم ح س}_1 \text{ ح س}_2}{n}$$

$$\sqrt{\sum (X_1 - \bar{X}_1)^2} = \sqrt{\sum (X_2 - \bar{X}_2)^2}$$

$$= 64.78$$

$$\sqrt{\sum (X_1 - \bar{X}_1)^2} = \sqrt{\sum (X_2 - \bar{X}_2)^2}$$

$$= 56.09$$

$$r = \frac{3347.90}{3633.51} = \frac{3.79 - 3351.69}{56.09 \times 64.78} = 0.921$$

أى أن هناك ارتباط طردى قوى بين كمية الأمطار الساقطة وشدة انسياب المياه فى النهر.

معامل ارتباط الرتب Rank Correlation Coefficient

فى حالة إذا كان توزيع المفردات لمتغيرين أو لأحدهما فى عينة غير معتدل، أو

عندما لا يكون مثل هذا التحديد متاح، فبدلاً من استخدام القيم الخاصة بهما فإنه يمكن ترتيب أو حجمها أو غير ذلك. وفي مثل هذه الحالة يستخدم مقياساً آخر لحساب نوع ودرجة الترابط بين المتغيرين يعرف باسم «معامل ارتباط الرتب». وهو أداة إحصائية غير باراميتريّة (غير معلمية) تقيس العلاقة بين نوعين من البيانات الترتيبية لمتغيرين. كما يعد معامل ارتباط الرتب الأداة البديلة لمعامل ارتباط ضرب العزوم. وذلك في حالة عدم افتراض التوزيع المعتدل لبيانات كل من المتغيرين. أو بعبارة أخرى لا يشترط في حساب معامل ارتباط الرتب أن يكون توزيع البيانات الأصلية للمتغيرين معتدلاً، ولكنه يشترط أن تكون بيانات المتغيرين من نوع بيانات الرتب.

ومن الطبيعي أن يكون هناك اختلاف بين قيمتي معامل الارتباط للقيم الأصلية (معامل ارتباط ضرب العزوم أو معامل ارتباط بيرسون) ومعامل الارتباط للرتب. والسبب في ذلك يرجع إلى استبدال القيم الأصلية للمفردات برتب خاصة، وفي هذه العملية بعض التقريب. كما أن معامل ارتباط الرتب أقل دقة من معامل ارتباط ضروب العزوم الذي يتأثر بأى تغير في القيم الأصلية التي تسجل عن مفردات العينة، بينما لا يتأثر معامل ارتباط الرتب بذلك لأن زيادة قيمة مفردة أو نقصها ولو بسيطاً لا يغير من وضع المفردة (ترتيبها) داخل العينة.

معامل ارتباط سبيرمان Spearman's Rank Correlation Coefficient

يشترط عند حساب معامل الارتباط بين متغيرين بطريقة سبيرمان أن لا يقل عدد المفردات (الحالات) المكونة للعينة عن عشرة مفردات. وتعمد طريقة حساب معامل ارتباط سبيرمان على إعطاء المفردات رتباً لتحل محل القيم العددية الأصلية، حسب الأهمية أو الحجم، لكل متغير من المتغيرين قيد التحليل. ويلزم حساب هذا المعامل أن نرتب القيم الزصلية للمفردات ترتيباً تصاعدياً أو تنازلياً، ثم نحسب الفروق بين الرتب لكل حالة من المتغيرين، ثم تربع هذه الفروق حتى نتخلص من الإشارات الحسائية. وبعد ذلك يمكن إيجاد قيمة معامل سبيرمان باستخدام الصيغة التالية:

$$\text{س} = ١ - \frac{٦ \text{ مج ف } ٢}{٣٧ - ٧} \dots\dots\dots (١١ - ٦)$$

أو:

$$\text{س} = ١ - \frac{٦ \text{ مج ف } ٢}{٢٧ - ١} \dots\dots\dots (١١ - ٧)$$

حيث (ف) هي لفروق بين رتبتى كل حالة، ن هي هدد أزواج الرتب.

مثال (٦):

نفرض أن هناك عشر مناطق صناعية رتبت ترتيباً تصاعدياً أو تنازلياً من ناحية إنتاج الآلات الميكانيكية وإنتاج السيارات وكانت الرتب على النحو التالى:

المنطقة	أ	ب	ج	د	هـ	و	ز	ح	ط	ك
الرتبة فى إنتاج الآلات الميكانيكية	٢	١	٤	٣	٥	٧	٦	٩	١٠	٨
الرتبة فى إنتاج السيارات	٢	٣	٥	١	٤	٩	٦	٧	٨	١٠

يلاحظ من البيانات أن المنطقة (ب) هي الأولى فى إنتاج الآلات الميكانيكية وثالثة فى إنتاج السيارات، بينما المنطقة (د) رتبته الأولى فى إنتاج السيارات والثالثة فى إنتاج الآلات الميكانيكية.

ويمكن الحصول على قيمة معامل ارتباط سبيرمان الرتب باستخدام القانون السابق كما يلى:

جدول رقم (١١-٧)
خطوات حساب معامل ارتباط سيرمان بين إنتاج
الآلات الميكانيكية وإنتاج السيارات

النقطة	رتبة إنتاج الآلات الميكانيكية	رتبة إنتاج السيارات	الفرقة (ف)	ف
أ	٢	٢	صفر	صفر
ب	١	٣	٢ -	٤
ج	٤	٥	١ -	١
د	٣	١	٢ +	٤
هـ	٥	٤	١ +	١
و	٧	٩	٢ -	٤
ز	٦	٦	صفر	صفر
ح	٩	٧	٢ +	٤
ط	١٠	٨	٢ +	٤
ك	٨	١٠	٢ -	٤
المجموع				٢٦

$$r_s = 1 - \frac{26 \times 6}{(1 - 100) 10} = 1 - \frac{156}{990}$$

$$= 1 - 0,157 = 0,843$$

وتشير قيمة معامل الارتباط (+0,843) في هذه الحالة إلى وجود علاقة
طردية قوية بين هذين المتغيرين.

مثال (٧):

نعود إلى المثال رقم (٢) الذي توضح بياناته مساحة حوض النهر وطوله لعينة من عشرة أنهار لتحسب منه معامل سبيرمان لارتباط الرتب:

جدول رقم (١١-٨)

حساب معامل ارتباط سبيرمان بين طول النهر ومساحة حوضه

٢		رتب ٢٣	رتب ١٣	(٢) مساحة النهر (كم ^٢)	(١) الطول الكلبي النهر (كم)
صفر	صفر	١ (أصغر مساحة)	١ (أقصر نهر)	٤,٠٨	٩,٤٠
١	١-	٢	٣	٤,٥٣	١٠,٨٥
١	١+	٣	٢	٤,٩٢	١٠,٦٩
صفر	صفر	٤	٤	٦,١٨	١٠,٩٧
صفر	صفر	٥	٥	٦,٢٢	١١,٩٤
٤	٢-	٦	٨	٦,٥٤	١٨,٦٩
١	١+	٧	٦	٧,٥٥	١٧,٢٣
١	١+	٨	٧	٨,٩٢	١٨,٤١
صفر	صفر	٩	٩	١٠,١٦	٢٠,١٤
صفر	صفر	١٠ (أكبر مساحة)	١٠ (أطول نهر)	١٢,٥٠	٢٠,٦٨
٨					الاجموع

ومن تطبيق قانون سبيرمان:

$$r_s = 1 - \frac{8 \times 6}{(1 - 100) 10}$$

$$0.9515 + = \frac{48}{990} - 1 =$$

أى أن قيمة معامل الارتباط بين المتغيرين هي $+0.9515$ وهى تدل على أن العلاقة بين المتغيرين علاقة طردية قوية، كما أنها قيمة تقترب كثيراً من قيمة معامل ارتباط بيرسون ($r = 0.902$) السابق حسابها لنفس البيانات.

مثال (٨):

البيانات الآتية تمثل معدلات النمو السكاني والنسبة المئوية لما يخص الفرد من الإنتاج القومي فى ١٤ دولة من دول العالم فى عام ١٩٧٠ والمطلوب حساب نوع ودرجة الارتباط بينهما باستخدام طريقة سبيرمان لارتباط الرتب.

جدول رقم (١١-٩)

حساب معامل ارتباط سبيرمان بين معدلات النمو السكاني والنسبة المئوية لما يخص الفرد من الإنتاج القومي في ١٤ دولة من دول العالم (عام ١٩٧٠)

الدولة	معدل النمو السكاني	الرتبة	النسبة المئوية من الإنتاج القومي للفرد	الرتبة	الفرقة (ف)	ف
البرازيل	٣,٠	٢	١,٦	١٠	٨-	٦٤,٠٠
نيجيريا	٢,٤	٤	٠,٣-	١٢,٥	٨,٥-	٧٢,٢٥
ألمانيا الغربية	١,٠	١١	٣,٤	٥,٥	٥,٥+	٣٠,٢٥
المملكة المتحدة	٠,٧	١٤	٢,٠	٨	٦+	٣٦,٠٠
إيطاليا	٠,٨	١٣	٤,٠	٣	١٠+	٥١٠٠,٠٠
فرنسا	١,١	٩,٥	٣,٧	٤	٥,٥+	٣٠,٢
المكسيك	٣,٥	١	٣,٤	٥,٥	٤,٥-	٢٠,٢٥
أسبانيا	٠,٩	١٢	٦,٥	١	١١+	١٢,٠٠
مصر	٢,٥	٣	١,١	١٠	٧-	٤٩,٠٠
بورما	٢,١	٦	١,٦	١٠	٤-	١٦,٠٠
يوغوسلافيا	١,١	٩,٥	٤,٢	٢	٧,٥+	٥٦,٢٥
أفغانستان	٢,٠	٧	٠,٣-	١٢,٥	٥,٥-	٣٠,٢٥
هولندا	١,٣	٨	٣,٠	٧	١+	١,٠٠
الجزائر	٢,٣	٥	٣,٥-	١٤	٩-	٨١,٠٠
المجموع						٧٠٧,٥٠٠

(المصدر: World Bank Atlas, 1970)

$$\frac{707.5 \times 6}{(1 - 196) 14} - 1 = \text{س}$$

$$1,000 - 1 = \frac{4240}{2730} - 1 =$$

$$0,000 - =$$

وتدل قيمة معامل الارتباط المحسوبة (-0,000) على وجود علاقة عكسية بين المتغيرين، بمعنى أنه في حالة زيادة النسبة المئوية لنمو السكان تقل النسبة المئوية لما يخص الفرد من الإنتاج القومي في هذه الدول. وإذا فحصنا البيانات في الجدول سنرى أن الأرقام التي تدل على ارتفاع معدل النمو السكاني وانخفاض نسبة ما يخص الفرد من لإنتاج القومي تختص بها الدول النامية (مثل المكسك والبرازيل)، يعكس ما هو ملاحظ على نفس المعدلات بالنسبة للدول المتقدمة (مثل المملكة المتحدة، وألمانيا الغربية).

وعلى الرغم من أن صيغة سبيرمان ليست بدقة صيغة معامل ارتباط العزوم، إلا أنها بسيطة في الاستخدام ويفضلها الكثير من الباحثين لأنها سهلة ولا تتطلب عمليات حسابية معقدة. ولكن هناك بعض العيوب التي تؤخذ على معامل سبيرمان لارتباط الرتب منها ما هو خاص بانعدام المعنى الطبيعي للمفرق بين رتبتين، ومعنى تربيع هذا الفرق، ومنها ما يتصل بالتوزيع الذي نحصل عليه من العينات المختلفة لحساب قيمة هذا المعامل.

معامل كندال لارتباط الرتب Kendall's Rank Correlation Coefficient

يفضل استخدام معامل كندال كثيراً عن معامل سبيرمان في قياس درجة الارتباط لأنه أسهل في حسابه، كما يمكن استخدامه في حالة العينات الصغيرة (عدد المفردات أقل من 10). ويتطلب حساب هذا المعامل إعادة ترتيب رتب مفردات أحد المتغيرين حسب الترتيب لاعادي إما تصاعدياً أو تنازلياً مع ترك رتب

المتغير الآخر بدون إعادة ترتيبها، ثم إيجاد قيمة معامل الارتباط بينهما بعد تطبيق معادلة كندال الآتية:

$$R_K = \frac{\text{مج د}}{n^{1/2} (1 - n)} \quad \text{أو} \quad \frac{2 \text{ مج د}}{n (1 - n)} \quad \dots\dots\dots (14-8)$$

حيث د هي مجموع الفروق بين الرتب. ويمكن الحصول على أكبر عدد من فروق الرتب إذا كانت كل الرتب في ترتيبها العادي، فإذا كان لدينا ن من الرتب فإن أكبر عدد ممكن من فروق الرتب هو $n^{1/2} (1 - n)$.

ويشبه معامل كندال لارتباط الرتب معامل سبيرمان السابق شرحه، من حيث أن قيمته تنحصر بين $+1,0$ ، $-1,0$. ونحصل على القيمة ± 1 إذا كانت الرتب المتناظرة متفقة تماماً ويسمى الارتباط في هذه الحالة بالارتباط التام الموجب. أما إذا كانت الرتب عكسية تماماً فإننا نحصل على قيمة لمعامل الارتباط تساوي $-1,0$ ، ويسمى الارتباط في هذه الحالة بالارتباط التام السالب.

وهناك طريقتان يمكن بواسطتهما حساب معامل كندال لارتباط الرتب، ونعرض فيمايلي خطوات الحساب التي تتبع في كل منهما.

الطريقة الأولى:

ينحصر حساب معامل كندال بهذه الطريقة في الخطوات الآتية التي نطبقها على المثال التالي:

مثال (٩):

لدراسة العلاقة الارتباطية بين كمية المطر والارتفاع عن سطح البحر عن طريق استخدام معامل كندال، أخذت البيانات الموضحة في الجدول التالي (جدول رقم ١٤-١) لستة من محطات الأرصاد الجوية في إحدى المناطق ذات الطبيعة الجبلية.

جدول رقم (١١-١٠)

طريقة حساب معامل كندال بين متغيري المطر وارتفاع الأرض عن سطح البحر لستة محطات جوية

المحطة	الارتفاع (س _١) (متر)	كمية المطر (س _٢) (مليمتر)	ترتيب (س _١)	ترتيب (س _٢)
أ	٢٥٠٠	٢٠٠	١	٢
ب	١٨٠٠	٢٢٥	٢	١
ج	٩٠٠	١٩٠	٣	٤
د	٦٠٠	٢١٠	٤	٣
هـ	٤٠٠	١٢٠	٥	٦
و	١٥٠	٨٥	٦	٥

١- ترتب المحطات ترتيباً عادياً تنازلياً (أو تصاعدياً) لمتغير واحد وهو في مثالنا، الارتفاع (س_١).

٢- نضع ترتيب المحطات بالنسبة للمتغير الثاني (كمية المطر س_٢) كما هي (أى بدون ترتيبها) مقابل ترتيب ارتفاعها، وذلك على النحو التالي:

المتغير (س _١)	١	٢	٣	٤	٥	٦
المتغير (س _٢)	٢	١	٤	٣	٦	٥

٣- نبدأ بملاحظة ترتيب (س_٢) غير المرتب ترتيباً عادياً بادئين من ترتيب البداية وهو (٢)، ثم نسجل لكل ترتيب اختلافه عن الترتيب الذى يليه بإعطاء القيمة + ١ للترتيب الأكبر، والقيمة - ١ للترتيب الأصغر. وفي هذه الحالة نجد أن الرتب التى تزيد عن الرتبة ٢ هي ٣، ٤، ٥، ٦ والتى تقل عنها هي ١ وتكون القيم المقابلة لها مجتمعة هي (-١، +١، +١، +١، +١، +١)

والمجموع الجبرى لها يساوى $3+$ (أى $(4+) + (1-) = 3+$) ، ولا يؤخذ هذا الترتيب فى الاعتبار مرة أخرى.

٤- وهكذا بالنسبة لباقى الرتب يمكن إيجاد القيمة المقابلة للرتب الأكبر والأصغر من الرتبة التالية للمتغير الثانى (س). وفى النهاية تقوم بجمع القيم المقابلة لجميع رتب هذا المتغير، فتحصل على المجموع الكلى لقيم الاختلاف، والتى يرمز لها بالرمز (د)، وذلك على النحو التالى:

القيم المقابلة	ترتيب (س)
$3+ = (1-) + (4+)$	٢
$4+ = (4)$	١
$1+ = (2+) + (1-)$	٤
$2+ = (2+)$	٣
$1+ = (1+)$	٥
صفر = (صفر)	٦
<u><u>$11+ =$</u></u>	المجموع (د)

٥- وبما أن معامل كندال لارتباط الرتب عبارة عن مقياس لاختلاف رتب أو لفروق ترتيب أحد المتغيرين عندما يكون المتغير الآخر مرتباً ترتيباً (أى فى تسلسل منتظم)، فإن قيمة المعامل يمكن الحصول عليها بقسمة المجموع العام للقيم المحسوب من اختلاف الرتب (د) على المجموع الأكبر المحتمل $1/2$ ن (١-ن) كما يلى:

$$K = \frac{\text{مجموع د}}{\frac{1}{2} \text{ ن (١-ن)}} = \frac{11}{\frac{1}{2} \text{ ن (١-٦)}} = \frac{11}{\frac{1}{2} \text{ ن (١-٦)}} = 0.733$$

٦- تدل قيمة معامل الارتباط (+ ٠,٧٣٣) على أن هناك علاقة طردية متوسطة بين متغيرى الارتفاع وكمية المطر.

مثال (١٠):

نعود إلى المثال رقم (٨) ونتخذ من بياناته الخاصة بمتغيرى معدل النمو السكاني والنسبة المئوية لما يخص الفرد من الإنتاج القومي في ١٤ دولة مختلفة من دول العالم أساساً لحساب معامل ارتباط الرتب ل Kendall.

١- نرتب الدول ترتيباً عادياً (تصاعدياً) بالنسبة لمتغير معدل النمو السكاني (س١).

٢- ننظم ترتيب الدول بالنسبة للمتغير الثاني (نصيب الفرد من الدخل القومي ش٢) مقابل ترتيب معدل نموها السكاني كمايلي:

(س١) ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩,٥ ١١ ١٢ ١٣ ١٤

(ش٢) ٨ ٣ ١ ٥,٥ ٤ ٢ ٧ ١٢,٥ ١٠ ١٤ ١٢,٥ ١٠ ٥,٥

٣- وكما فعلنا في المثال السابق نبدأ بملاحظة ترتيب المتغير (س٢) الذى لم يترتب ترتيباً عادياً، ونسجل لكل ترتيب اختلافه عن الترتيب الذى يليه بإعطاء القيمة (+١) للرتب الأكبر، والقيمة (-١) للرتب الأصغر، ثم نجمع الناتج مبتدئين بالترتيب ٥,٥ الذى هو أول ترتيب للمتغير (س٢). ونظراً لوجود بعض الرتب المتكافئة (المتساوية فى الترتيب) فإننا نعطى الرتب التى تقع على يسار الرتبة المتساوية معها فى الترتيب القيمة (صفر). فمثلاً الرتبة الأولى ٥,٥ تقع على يسارها الرتب الأكبر والأصغر منها وبينها رتبة أخرى مساوية لها هى ٥,٥ وتعطى للأخيرة القيمة (صفر). كذلك نلاحظ أن الربتين ٩,٥، ٩,٥ للمتغير (س١) تقابلان من رتب المتغير الثاني (س٢) الرتبة ٣ والرتبة ٤، وفى هذه الحالة فإن موضع رتب المتغير الثاني تعتمد اعتماداً كبيراً على موضع الرتب المتكافئة للمتغير الأول، ولكن يصعب تحديد أى من رتبى المتغير الأول يجب أن تقابل الرتبة ٢ أو تقابل الرتبة ٤ من رتب المتغير الثاني. وهذا التحديد

له أهمية كبيرة فى تحديد المجموع الكلى (د) للاختلاف بين الرتب. فمثلاً إذا وضعنا الرتبة ٤ قبل الرتبة ٢، ليزاد المجموع الكلى (د) ٢. وللتغلب على هذه المشكلة تبقى الرتبة الأولى من المتغير (س_٢) التى تقابل إحدى الرتب المتكافئة من رتب المتغير (س_١) كما هى فى وضعها وبنفس ترتيبها، بينما يعطى للرتبة الثانية من المتغير (س_٢) التى تقابل الرتبة المتكافئة الأخرى من رتب المتغير (س) القيمة صفر. فمثلاً إذا كانت الرتبة ٢ هى الرتبة الأولى من رتب (س_٢) والتى سنقابل الرتبة المتكافئة الأولى من رتب (س_١) فإن الرتبة ٤ تأخذ القيمة (صفر) عند حساب المجموع الكلى لاختلاف الرتب (د). وهكذا إذا كانت الرتبة ٤ هى الأولى فإن الرتبة ٢ من رتب (س_٢) تأخذ القيمة (صفر)، وذلك لأنه فى كل من الحالتين يكون (١+) + (١-) = صفر.

٤- يحسب المجموع الكلى للفروق بين رتب س_١، س_٢ السابقة على النحو التالى:

$$(د) = (٧-) + (٧-١) + (٩-) + (٨-١) + (٧-٣) + (٧-٣) + (٤-٨) =$$

$$٣٣- = (١+) + (٢+) + (٢-١) + (٢-٢) + (١-٣) + (٥-١)$$

وبالتعويض عن (د) بالقيمة (٣٣-) ون = ١٤ فى معادلة كندال لحساب قيمة معامل ارتباط الرتب، نجد أن:

$$r_k = \frac{D}{\frac{1}{2} n(n-1)}$$

$$٠,٣٦٣- = \frac{٣٣-}{\frac{1}{2} (١٤)(١٤-١)} =$$

وكما هو واضح فإن قيمة معامل ارتباط كندال المحسوبة (٠,٣٦٣-) أقل بكثير من معامل ارتباط الرتب السابق لنفس البيانات.

الطريقة الثانية:

نفرض - على سبيل المثال - أننا نريد حساب معامل كندال لارتباط الرتب للمثال رقم (٢) الخاص بطول النهر (س_١) ومساحة حوضه (س_٢) فنتبع الخطوات التالية.

١ - نقوم بإعادة ترتيب رتب أحد المتغيرين حسب الترتيب العادي (١، ٢، ٣، ...) فمن المثال السابق الذى يبين طول النهر (س_١) ومساحة حوضه (س_٢) نحصل على:

س_١: ١٠، ٨٥، ٩، ٤٠، ١٠، ٦٩، ١٠، ٩٧، ١١، ٩٤، ١٨، ٦٩، ١٧، ٢٣، ١٨، ٤١، ١٤، ٢٠، ٦٨، ٢٠

س_٢: ٤، ٠٨، ٤، ٥٣، ٤، ٩٢، ٦، ١٨، ٦، ٢٢، ٦، ٥٤، ٧، ٥٥، ٨، ٠٢، ١٠، ١٦، ١٢، ٥٠

ترتيب س_١: ١ ٣ ٢ ٤ ٥ ٨ ٦ ٧ ٩ ١٠

ترتيب س_٢: ١ ٢ ٣ ٤ ٥ ٦ ٧ ٨ ٩ ١٠

٢ - نبدأ بملاحظة المتغير الذى لم يترتب ترتيباً عادياً (المتغير س_١) بادئين بالرتبة الأولى (وهى ١) ونبحث عن الرتبة العادية المناظرة لها فى المتغير س_٢ فنجدها ١ وهذا الترتيب على يساره ٩ رتب ولا توجد رتبة أعلى يعينه، وعلى ذلك يكون فرق عدد الرتب = (٩ - صفر) = ٩+، ثم نشطب الرتبة ١ من الرتب العادية فنحصل على:

٣ ٤ ٥ ٨ ٦ ٧ ٩ ١٠

٢ ٤ ٥ ٦ ٧ ٨ ٩ ١٠

ونبحث عن الترتيب المناظر للرتبة ٢ فى المتغير (س_١) فنجدها ٣، ونجد على يسارها ٧ رتب وعلى يمينها توجد رتبة واحدة وبذلك يكون فرق عدد الرتب = (٧ - ١) = ٦+، ثم نشطب ٢ من الرتب العادية فنحصل على:

٣ ٤ ٥ ٨ ٦ ٧ ٩ ١٠

٢ ٤ ٥ ٦ ٧ ٨ ٩ ١٠

فنجـد أن ٢ (وهو الترتيب المناظر للترتيب ٣) يقع على يساره ٧ رتب وعلى يمينه لا توجد أية رتب، فيكون الفرق في عدد الرتب هو $7- = 0$ ، وهكذا، فتكون الفروق في عدد الرتب هي:

$$+9, +6, +7, +6, +5, +2, +1, +2, +1, +0, -1, -2, -3$$

ويكون مجموع الفروق في عدد الرتب = ٣٩

٣- نحسب معامل كندال لارتباط الرتب من المعادلة:

$$R_k = \frac{2 \text{ مجـد د}}{n(n-1)}$$

$$= \frac{78}{90} = \frac{29 \times 2}{(1-10) 10} = 0.886 + =$$

وكما هو واضح فهو قريب من معامل ارتباط الرتب السابق لسييرمان.

اختبار المعنوية الاحصائية للارتباط:

سبق أن قلنا أن تحليل الارتباط ماهو إلا وصف إحصائي لدرجة ترابط وعلاقة المتغيرين قيد التحليل، كما ذكرنا أن قيمة معامل الارتباط المحسوبة بإحدى الطرق الإحصائية المختلفة لاتدل دلالة أكيدة على وجود علاقة بين المتغيرين، إذ قد تكون النتائج التي نحصل عليها متأثرة بعامل الصدفة الناتج من خطأ في أسلوب المعالجة. وعلى ذلك يجب عمل اختبار لقيمة معامل الارتباط للتأكد به من درجة احتمال أن الارتباط لا يحدث بطريق الصدفة، أو بمعنى آخر نتعرف به على مدى معنوية هذا الارتباط وتأثير حجم العينة أو البيانات موضع البحث والتحليل.

ولاختبار معاملات الارتباط السابقة يوضع فرض العدم القائل أن قيمة الارتباط بين المتغيرين هي صفر، أو بعبارة أخرى لا يوجد ارتباط بين المتغيرين (أي أنهما مستقلين عن بعضهما البعض). ولاختبار هذا الفرض فالتا نعين مستوى الدلالة أو

المعنوية Significance Level المطلوب سواء لمستوى احتمال ٠,٠٥ أو ٠,٠١ ثم نقوم بمقارنة قيمة معامل الارتباط المحسوبة بالقيم النظرية فى الجداول الخاصة بكل نوع من أنواع معاملات الارتباط الثلاثة (راجع ملاحق الجداول الاحصائية بنهاية الكتاب). فإذا كانت قيمة معامل الارتباط المحسوبة أكبر من نظيرتها فى الجدول بدرجة حرية مساوية لعدد أزواج القيم المشاهدة مطروحاً منها ٢ فإن هذا يعنى رفض فرض العدم وقبول الفرض البديل وهو أنه يوجد فعلاً ارتباط بين المتغيرين (أى أن الارتباط له دلالة إحصائية عند مستوى المعنوية المطلوب). وسنوضح فيمايلي كيفية اختبار كل نوع من معاملات الارتباط على حدة.

أولاً: بالنسبة لمعامل ارتباط بيرسون يستخدم توزيع ستيودنت - ت لاختبار قيمة معامل الارتباط المحسوبة، وذلك بعد تعديل الصيغة إلى:

$$t = \frac{r \sqrt{2-n}}{\sqrt{1-r^2}} \quad \text{أو} \quad r = \frac{\sqrt{2-n}}{\sqrt{2-n+r^2}} \quad \dots\dots\dots (١١-٩)$$

حيث r هى قيمة معامل ارتباط بيرسون، n هى عدد أزواج القيم فى المثال رقم (٢) الذى عرضناه عن طول النهر ومساحة حوضه وجدنا أن $r = ٠,٩٠٢$ وأن $n = ١٠$ وبالتعويض فى الصيغة السابقة نحصل على

$$t = \frac{0,902 \sqrt{2-10}}{\sqrt{1-0,902^2}}$$

وحيث أن قيمة t النظرية بدرجات الحرية $2-10 = ٨$ هى ٢,٣١ لمستوى دلالة ٠,٠٥ أصغر من قيمة t المحسوبة، فإن ذلك يعنى أننا نرفض فرض العدم القائل أنه لا يوجد اختلاف بين القيمة المحسوبة لمعامل الارتباط والصفر (الذى يعنى عدم وجود علاقة بين المتغيرين)، ونقبل الفرض البديل وبمعنى آخر أن هناك اختلاف جوهري بين معامل الارتباط المحسوب ومعامل ارتباط صفر، وأن قيمة معامل الارتباط المحسوبة لها دلالة (معنوية) إحصائية عند مستوى ٠,٠٥.

وفى المثال رقم (٥) الذى يختص بحساب قيمة معامل ارتباط بيرسون بين كمية الأمطار وشدة سريان المياه فى أحد الأنهار فإنه بالتعويض فى الصيغة السابقة نحصل على:

$$t = 0.921 = \frac{\sqrt{2 - 16}}{2(0.921) - 1} = 8.846$$

وحيث أن قيمة ت النظرية بدرجات الحرية $2 - 16 = 14$ هى ٢,١٥ لمستوى دلالة ٠,٠٥ أصغر من قيمة ت المحسوبة، فإن ذلك يعنى أن معامل الارتباط المحسوب له دلالة (معنوية) إحصائية عند مستوى ٠,٠٥.

ثانياً: فيما يختص بمعامل سبيرمان لارتباط الرتب فإنه يمكن استخدام بيانات الجدول الاحصائى الخاص به (راجع الجداول الإحصائية بنهاية الكتاب) لاستخلاص القيمة المتوقعة لمعامل الارتباط حسب حجم العينة (ن) ومستوى الدلالة المطلوب. وفى حالة المثال رقم (٨) الخاص بحساب الارتباط بين معدل النمو السكانى ونصيب الفرد من الإنتاج القومى، نجد أن حجم العينة (عدد الدول) ١٤، وأن مستوى الدلالة هو ٠,٠٥ (أى احتمال أن ٥٪ من الحالات تكون فيها قيمة معامل الارتباط راجعة إلى الصدفة)، والقيمة المتوقعة التى نحصل عليها من الجدول الاحصائى هى ٠,٤٥٦ وهى قيمة تقل عن القيمة التى حصلنا عليها فى مثالنا وهى -٠,٥٥٥ أى أننا نستطيع أن ننفى بأن لقيمة التى حصلنا عليها ترجع إلى الصدفة بمستوى الدلالة المطلوب، أو بعبارة أخرى أن هناك احتمال مقداره ٩٥٪ فإن لا تكون قيمة معامل الارتباط التى حصلنا عليها قد حدثت بفعل الصدفة أو العشوائية فى التوزيع، وأن الارتباط (العلاقة) بين المتغيرين له دلالة إحصائية عند مستوى المعنوية المطلوب.

كما يمكن استخدام الصيغة الاحصائية المعدلة لاختبار ستيودنت - ت لمعايرة قيمة معامل ارتباط سبيرمان المحسوبة لتحديد دلالة أو معنوية هذه القيمة بنفس الطريقة التى اتبعتها عند معايرة قيمة ارتباط بيرسون.

ثالثاً: أما في حالة معامل ارتباط كندال فإن التعرف على دلالة أو معنوية قيمة المعامل يختلف حسب عدد المفردات. فإذا كان عدد المفردات ١٠ فإننا نلجأ إلى الجداول الإحصائية الخاصة بهذا المعامل (انظر ملحق الجداول الإحصائية) والتي توضح درجة الاحتمال المرتبطة بعدد مفردات العينة (ن). ففي المثال رقم (٩) الخاص بمعرفة درجة الترابط بين كمية المطر والارتفاع عن سطح البحر كان عدد مفردات العينة ٦ والمجموع الكلي للفروق بين هذين المتغيرين (د) هو ١١ فتصبح درجة الاحتمال ٠,٠٢٨ (٣٪ تقريباً) أى أن هناك ٣٪ من الحالات يكون فيها الارتباط قد حدث بفعل الصدفة، وبعبارة أخرى أن قيمة معامل الارتباط التي حصلنا عليها وهي ٠,٧٣٣ ذات دلالة إحصائية عند مستوى معنوية ٠,٠٥.

أما إذا كان عدد المفردات للمتغيرين أكثر من ١٠، كما هي الحال في المثال رقم (١٠) الخاص بمعرفة درجة الترابط بين معدل النمو السكاني ونصيب الفرد من الإنتاج القومي في بعض دول العالم، فإننا نستخدم توزيع (ز) لمعايرة قيمة معامل الارتباط المحسوبة لتحديد دلالة أو معنوية هذه القيمة، ويتم ذلك بالمعادلة الآتية:

$$(ز) = \frac{\sqrt{\frac{2(2 + 5)}{(1 - 0.733)^2}}}{\sqrt{11}} \quad \dots\dots\dots (١١-١٠)$$

وتحدد درجات الحرية على أساس أنها تساوى (ن - ٢).

ويمكن حساب قيمة (ز) لمعامل كندال لارتباط الرتب الذي حصلنا عليه من المثال رقم (١٠) كما يلي:

$$(ز) = \frac{\sqrt{\frac{2(2 + 33)}{13 \times 126}}}{\sqrt{1638}} = \frac{0.363 -}{0.362 -} = 1.81 -$$

وبالرجوع إلى جداول توزيع (ر) لنظريه نجد أن قيمة (ز) المحسوبة وهى ١,٨١ لها احتمال ٠,٣٦ أى أن قيمة معام كندال المحسوبة لا ترجع إلى الصدفة فى مستوى معنوية ٠,٠٥ وهذا يعنى بشكل آخر أن هناك احتمال مقداره ٩٥٪ بأن لا يكون الارتباط فى المثال السابق قد حدث بفعل الصدفة، لأن عامل الصدفة يتدخل فى ٣٦٪ من الحالات. وعلى العموم فإنه مع معامل ارتباط كندال نجد أن الاحتمالات الخاصة بقييم د (الفرق بين الرتب) المحسوبة لعدد المفردات (ن) التى تتراوح بين ١٠, ٤٠ تقل كلما زادت قيمة د، وبالتالى يتخذ ذلك دليلاً على صغر فرصة حدوث الارتباط بالصدفة.

الفصل الثاني عشر
تحليل الانحدار
Regression Analysis

تحليل الانحدار

Regression Analysis

ذكرنا في الفصل السابق عن تحليل الارتباط أن الهدف من حساب معامل الارتباط هو معرفه درجه العلاقه أو مقدار الترابط بين متغيرين (ظاهرتين). إلا أنه إذا ما وجدت علاقة بين متغيرين فإننا ربما نحتاج إلى التوقع (أو التنبؤ prediction) بسلوك أحد المتغيرين في ضوء تأثيره بمتغير آخر أو بعدة متغيرات أخرى ، أو إذا كانت هناك رغبة في تقدير مدى تأثير كل متغير من المتغيرات على متغير آخر. وواضح أن مثل التقدير يزداد دقة كلما كان الارتباط شديداً. ويسمى المتغير الذي يراد دراسة سلوكه ومعرفة مدى تأثيره بالمتغيرات الأخرى المتغير التابع (ص) Dependent Variable ويطلق على المتغير الذي يؤثر في سلوك المتغير التابع بالمتغير المستقل Independent Variable. وفي بعض الأحيان يسمى المتغير التابع باسم المتغير المتنبأ له Predictand Variable أو المتغير المعيار Criterion Variable ، كما أنه يطلق على المتغير المستقل اسم المتغير المتنبئ به Predictor Variable أو المتغير المفسر Explanatory Variable .

كما وقد سبق أن أوضحنا أنه لا يشترط لدراسة الارتباط بين متغيرين أن تكون هناك علاقة دالية بينهما ، ولكن إذا كانت العلاقة الدالية بين متغيرين يمكن وصفها إحصائياً بخط مستقيم سميت هذه العلاقة «بعلاقة انحدار خطية» linear Regression. وبذلك يختص الانحدار (أو ما يعرف أحيانا بالارتداد أو الاعتماد)

بدراسة العلاقة بين متغيرين على هيئة علاقة دالية بحيث يمكن التنبؤ منها عن أحد المتغيرين بمعلومية المتغير الآخر.

أهداف تحليل الانحدار:

يستخدم تحليل الانحدار كأسلوب احصائي كمي في النواحي التالية :

١- تقدير العلاقة بين متغيرين على شكل علاقة دالية: $ص = (س) أو س = (ص)$ والتي عن طريقها يمكن معرفة التغير في أحد المتغير على أساس تأثيره بالمتغير الآخر. أو بعبارة أخرى توقع وتنبؤ سلوك المتغير التابع في ضوء تأثيره بالمتغير أو المتغيرات المستقلة.

٢- قياس مدى الارتباط الكلى بين المتغير التابع والمتغير أو المتغيرات المستقلة.

٣- تقدير نسبة تفسير كل متغير مستقل للاختلاف في المتغير التابع.

٤- إجراء سلسلة من الاختبارات الفرضية لأى من العلاقات الثلاثة السابقة.

أنواع تحليل الانحدار:

هناك ثلاثة أنواع رئيسية لتحليل الانحدار نجعلها فيما يلى :

١- الانحدار البسيط Simple Regression و هو يستخدم لدراسة العلاقة بين متغيرين فقط.

٢- الانحدار الجزئى Partial Regrenion و هو يدرس العلاقة بين المتغير التابع وواحد فقط من المتغيرات المستقلة بفرض أن للعوامل الأخرى ثابتة (أى بإهمال تأثير العوامل الأخرى).

٣- الانحدار المتعدد Multiple Regression وهو يحدد مقدار العلاقة بين المتغير التابع والمتغيرات المستقلة كلها. وهناك نوع آخر مشابه للانحدار المتعدد يسمى بالانحدار التدرجى Stepwise Regression وهو يعطى نسبة تفسير كل متغير مستقل فى اختلاف المتغير التابع، وتكون نسبة التفسير مرتبة حسب أهمية

كل متغير مستقل (أى أنه يبدأ بتحديد أعلى نسبة أو أهم متغير وينتهى بأقل نسبة أو المتغير أقل أهميه فى تفسير الاختلاف الذى يحدث فى المتغير التابع).

ورغم وجود بعض الاختلافات فى العمليات الحسابية لكل نوع من أنواع التحليل السابقة إلا أنها تسير فى نفس الاتجاه تقريباً، وهو تحديد معادلة خط الانحدار للعلاقة بين متغيرين أو عدة متغيرات. وسنقتصر فى دراستنا فى هذا الفصل على تحليل النوع الأول من الانحدار (تحليل الانحدار البسيط) كأحد التحليلات الشائعة الاستخدام، حيث أن النوعين الآخرين (الانحدار الجزئى والانحدار المتعدد) يحتاجان إلى جهد كبير ووقت طويل فى عملياتهما الحسابية كلما زاد عدد الحالات Cases والمتغيرات Variables. وقد ظهرت أهميه الحاسب الآلى Computer كعامل مساعد هام للقيام بمثل هذه العمليات الحسابية الطويلة. وكان لتقدم الدراسات الخاصة بنظم معالجة البيانات أن تعددت البرامج الجاهزة التى تستخدم فى مثل هذا النوع من التحليل الإحصائى. ومن أمثله هذه البرامج برنامج معروف فى مراكز الحاسبات الآلية الكبيرة فى جامعات المملكة المتحدة ضمن مجموعة من البرامج يطلق عليها (S.P.S.S). ويقوم هذا البرنامج بحساب العلاقات بين عدد من المتغيرات المستقلة و المتغير التابع.

تحليل الانحدار البسيط :

تجدر الإشارة قبل الخوض فى تحليل الانحدار البسيط (بين متغيرين). أن نذكر أن هناك شرطاً أساسياً فى بيانات المتغيرات المستخدمة للتحليل وهو أن تكون هذه البيانات من نوع بيانات الفترة Intervals-Scaled data فى حالة المتغير التابع، ويستحسن أن تكون كذلك للمتغير المستقل. ولكن أحياناً يمكن أن تكون بيانات المتغير المستقل من نوع البيانات النوعية Nominally-Scaled data. كما لا بد أن نوضح أن هناك بون شاسع بين تحليل الارتباط وتحليل الانحدار. فعلى الرغم من تشابه العلاقات الرياضية بين الارتباط والانحدار إلا أنهما يختلفان عن بعضهما فى النواحي التالية:

١ - تحليل الارتباط عبارة عن مقياس وصفى بينما تحليل الانحدار مقياس كمي.

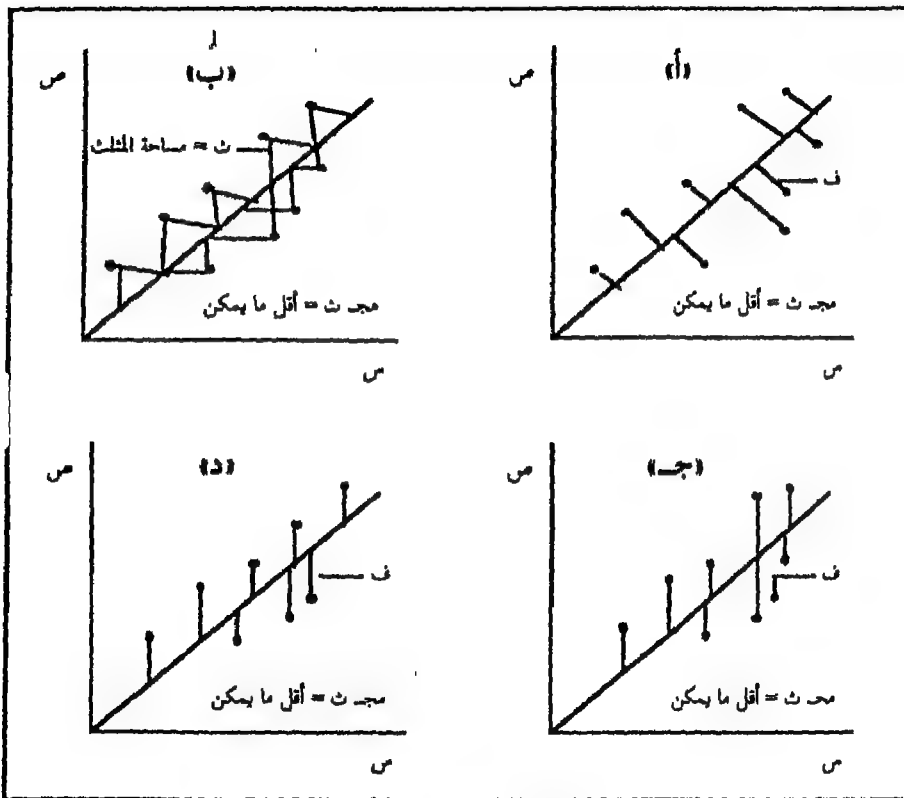
٢ - يشترط في تحليل الارتباط أن يكون توزيع بيانات كل المتغيرات توزيعاً معتدلاً، بينما يشترط في تحليل الانحدار أن تكون بيانات المتغير التابع (ص) فقط ذات توزيعاً معتدلاً، أما المتغير المستقل (س) فيجب أن تكون قيم مفرداته ثابتة Fixed (أي أن قيم (س) تقاس بدون أخطاء)، ولو أن الانحدار يتأثر بوحدات القياس - إلا أنه في بعض الحالات التي يحسب فيها الارتباط والانحدار فإننا نتغاضى عن الشرط الخاص بأن قيم المتغير المستقل (س) تكون قيماً ثابتة.

٣ - يشترط في تحليل الانحدار أن تكون هناك علاقة دالية بين المتغيرات، بينما لا يشترط ذلك في تحليل الارتباط.

و بصفة عامة فإن استخدام أسلوب الارتباط لا يحقق سوى قياس درجة العلاقة الارتباطية بين المتغيرين، بينما الهدف من أسلوب الانحدار هو دراسة التوقع أو التنبؤ بتغير المتغير التابع في ضوء معرفة التغيرات في المتغير المستقل وهو ما يطلق عليه بالعلاقة الدالية.

و يعتمد تحليل الانحدار لدراسة العلاقة بين ظاهرتين على تكوين فكرة مبدئية عن نوع هذه العلاقة وقوتها وذلك باستخدام ما يعرف بشكل الانتشار Scatter Diagram. فإذا مثلنا أزواج المشاهدات (القيم) الخاصة بالظاهرتين بيانياً نحصل على عدد من النقاط التي قد تقع تماماً على خط مستقيم فيكون الارتباط تاماً - أو قد تنحرف عنه أو تأخذ شكلاً آخر غير الخط المستقيم، وعلى العموم إذا كانت هناك علاقة تربط الظاهرتين فإن النقاط تنتشر بشكل منتظم حسب نوع العلاقة الموجودة (علاقة عكسية أو علاقة طردية). أما إذا كانت النقاط مبعثرة دون نظام ملحوظ فإن العلاقة تكون ضعيفة جداً أو منعدمة (أنظر الفصل السابق عن تحليل الارتباط). والخط الذي تنتشر حوله النقاط بانتظام يسمى خط الانتشار أو خط الانحدار (ويكون هذا الخط مستقيماً أو منحنيًا).

ومن المعلوم أن الخط الذي نوقفه لا يمر بجميع النقاط (إلا في حالات خاصة) في شكل الإنتشار وعلى ذلك تكون هناك بعض النقاط التي تنحرف عن هذا الخط، وبالتالي إذا اخترنا أى قيمة للمتغير المستقل بمعلومية احداثيها الأفقى و قدرنا قيمه الاحداثى الرأسى للمتغير التابع المناظرة لها فإن القيمة الأخيرة (المقدر) سوف تختلف عن قيمة الاحداثى الرأسى المشاهدة والفرق بين القيمتين المقدره والمشاهدة يسمى إنحراف النقطة (البعد الرأسى لها) عن الإنحدار (شكل رقم ١٢-١)



شكل رقم (١٢-١)

انحراف نقط تمثيل المتغيرين س ، ص عن خط الانحدار

ويمكن حساب معادلة خط الإنحدار بطريقة المربعات الصغرى التي من

خصائصها أن يكون مجموع مربعات انحرافات النقط عن خط الانحدار أصغر ما يمكن. ويكون خط الانحدار مستقيماً إذا كانت العلاقة بين المتغيرين علاقة دالية. والمعروف أن معادلة الخط المستقيم هي معادلة من الدرجة الأولى على صورة:

$$ص = م س \pm ج$$

حيث ص في هذه الحالة هي قيمة المتغير التابع، س هي قيمة المتغير المستقل، وحيث م مقدار ثابت يمثل ميل خط الانحدار على المحور الافقى، ويسمى معامل الانحدار Regression Coefficient، و ج مقدار ثابت أيضاً هو طول الجزء المقطوع من المحور الرأسى بواسطة خط الانحدار. وبتعين م، ج يتعين خط الانحدار كما يمكن تقدير قيمة (ص) وهي القيمة التي تمثل التوقع أو التنبؤ المطلوب حيث أن قيمة (س) معروفة.

ولإيجاد معادلة خط الانحدار على الصورة السابقة تحسب قيم م، ج التي تحقق الشرط العام لهذا الخط وهو أن مجموع مربعات انحرافات الأبعاد الرأسية للنقط عنه تكون أصغر ما يمكن. ويمكن الحصول عليها بواسطة عدة طرق أكثرها شيوعاً الطرق التالية:

$$(1) م = \frac{\frac{\sum (ص \times س)}{ن} - \frac{\sum ص \times \sum س}{ن^2}}{\frac{\sum (س^2)}{ن} - \frac{(\sum س)^2}{ن^2}}$$

$$(2) ج = \frac{\sum ص - م \sum س}{ن}$$

حيث س المتوسط الحسابى للمتغير المستقل، ص هي المتوسط الحسابى للمتغير التابع، ع س هي تباين المتغير المستقل.

$$(٣) م = \frac{\text{محد } س - \text{ن } س - \text{محد } ص}{\text{محد } س - \text{ن } س}$$

حيث أن هي عدد أزواج القيم للمتغيرين.

$$(٤) م = \frac{\text{محد } (س - ص) (س - ص)}{\text{محد } (س - ص)}$$

إلا أنه لتسهيل العمليات الحسابية فإنه يفضل استخدام الصيغة رقم (٣). أما قيمة (جـ) فتحسب بالطريقة التالية بعد معرفة قيمة (م).

$$\text{جـ} = \text{ص} - \text{م } س$$

وعن طريق معرفة قيمة كل من م، جـ يمكن تحديد قيم المتغير التابع (ص) في ضوء أى تغير في قيم المتغير المستقل (س). ويرمز لقيم (ص) المتوقعة بالرمز ص̂. وتكون المعادلة المستخدمة في التوقع أو التنبؤ الصحيح في حالة متغيرين فقط كالآتي:

$$\text{ص} = \text{م } س + \text{جـ}$$

مثال:

في المثال الخاص لدراسة العلاقة بين مساحة حوض النهر وطوله (مثال رقم: ٢ - الفصل الرابع عشر) فإنه لإيجاد المعادلة الخطية العلاقة بين هذين المتغيرين نقوم بوضع البيانات في جدول ونحسب منه قيمة (ص) وذلك على النحو التالي:

جدول رقم (١٢ - ١)
مساحة حوض النهر وطوله

س ^٢	س ^٢	المتغير التابع ص (مساحة حوض النهر)	المتغير المستقل س (طول النهر الكلي)
٣٨,٤	٨٨,٤	٤,٠٨	٩,٤٠
٤٩,٢	١١٧,٧	٤,٥٣	١٠,٨٥
٥٢,٦	١١٤,٣	٤,٩٢	١٠,٦٩
٦٧,٦	١٢٠,٣	٦,١٨	١٠,٩٧
٧٤,٣	١٤٢,٦	٦,٢٢	١١,٩٤
١٢٢,٢	٣٤٩,٣	٦,٥٤	١٨,٦٩
١٣٠,١	٢٩٦,٩	٧,٥٥	١٧,٢٣
١٦٤,٢	٣٣٩,٩	٨,٩٢	١٨,٤١
٢٠٤,٦	٤٠٥,٦	١٠,١٦	٢٠,١٤
٢٥٨,٥	٤٢٧,٧	١٢,٥٠	٢٠,٦٨
١١٦١,٩	٢٤٠٢,٧	٧١,٦	١٤٩,٠٠

$$س = ١٤,٩ \quad ص = ٧,١٦$$

وتطبيق الصيغة رقم (٣) لحساب قيمة (م) فإن:

$$= \frac{١٤,٩ \times ٧,١٦ \times ١٠ - ١١٦١,٩}{٢(١٤,٩) ١٠ - ٢٤٠٢,٧}$$

$$= \frac{٩٥,٠٦}{١٨٢,٦} = ٠,٥٢ +$$

قيمة جـ (ص - م س) فإن:

$$جـ = ٧,١٦ - ١٤,٩ \times ٠,٥٢ = -٠,٥٨٨$$

ويمكن حساب قيمة (م) على أساس الصيغة رقم (١)

$$\frac{\frac{149,98 \times 71,6}{10} - 11619}{\frac{2(149,98)}{10} - 2402,7} =$$

$$= \frac{95,06}{182,6} = 0,52 \text{ (وهي نفس النتيجة التي حصلنا عليها سابقاً)}$$

ويوضح الشكل رقم (١٢ - ٢) كيفية رسم خط الانحدار الذي يمثل العلاقة المتوقعة بين الظاهرتين مساحة حوض النهر وطوله. فبعد أن يتم رسم المحور الأفقى والمحور الرأسى توقع النقط التى تمثل قيم الظاهرتين (المتغيرين س، ص) على الرسم. فمثلاً النقطة (١) فى الشكل تبين نقطة التقاء قيمة ش المشاهدة (٩,٤٠) وقيمة (ص) المشاهدة (٤,٠٨). وتمثل النقطة (٥) فى الرسم نقطة التقاء س المشاهدة (١١,٩٤) و (ص) المشاهدة (٦,٢٢)، كذلك تمثل النقطة (٧) التقاء قيمة (س) المشاهدة (١٧,٢١) بقيمة (ص) المشاهدة (٧,٥٥) وهكذا. وبعد أن يتم تمثيل جميع النقط يرسم خط الانحدار وذلك بتوصيل نقطتين: الأولى تمثل قيمة (جـ) على محور الصادات وهى فى حالة المثال السابق (٠,٥٨٨-)، والنقطة الثانية تمثل نقطة التقاء قيمتى متوسط س، ص (س، ص) وهى فى هذه الحالة (١٤,٩٠)، (٧,١٦) وتمثلها النقطة (أ) فى الشكل. وبعد أن توصل هاتين النقطتين بخط يكتب على امتداده المعادلة التى يستدل بها عليه وهى عبارة عن:

$$\text{ص} = 0,52 \text{ س} - 0,588$$

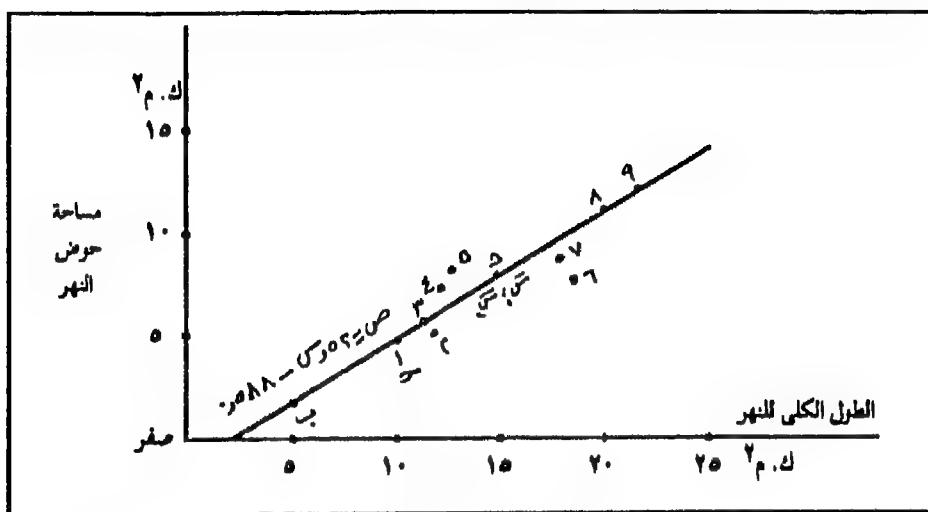
وهناك طريقة أخرى لرسم خط الانحدار بين النقط الموزعة على شكل الانتشار. وهى أن نطبق معادلة خط الانحدار التى حصلنا عليها وذلك لأية قيمتين من قيم (س)، وعن طريقها تحدد قيمتى (ص) المناظرتين وبعد توزيعهما على

الرسم توصل بينهما بخط ونمده حتى يلتقى بالمحور الرأسى. ونوضح ذلك عن طريق حساب قيمة (ص) عندما تكون قيمة (س) ١٥,٥ كما يلي:

$$(١) \text{ ص} = ٠,٥٨٨ - (٥ \times ٠,٥٢) = ٢,٠١$$

$$(٢) \text{ ص} = ٠,٥٨٨ - (٥ \times ٠,٥٢) = ٧,١٢$$

وتمثل النقطتان ب، د فى الشكل قيمتى (ص). ٢,٠١، ٧,١٢ المحسوبة لقيمتى (س) ١٥,٥ على الترتيب (لاحظ أن كل من القيمتين تقع على نفس خط الانحدار المرسوم بالطريقة الأخرى).



شكل رقم (١٢ - ٢)

شكل الانتشار وخط الانحدار للعلاقة بين مساحة حوض النهر وطوله

ومن الأهمية فى تحليل الانحدار أن نوضح ما إذا كان معامل الانحدار (م) له دلالة إحصائية تمكن الباحث من معرفة مدى ترابط العلاقة بين المتغيرين (س، ص)، أو بمعنى آخر توضيح ما إذا كان الارتباط بين المتغيرين قد تم بتأثير الصدفة. واختبار مستوى دلالة معامل الانحدار يعتمد على حساب قيمة (ت) حيث يكون الفرض المختبر هو أن معامل الانحدار (م) يساوى صفراً.

ومن المعروف فى تحليل الانحدار أنه يوجد نوعان من الانحرافات: النوع الأول يسمى الانحرافات عن الانحدار. وهى عبارة عن الفرق بين متوسط قيم المتغير التابع أو (ص) والقيم المتوقعة (ص̂) لهذا المتغير والمحسوبة من معادلة الانحدار، أما النوع الآخر من الانحرافات فيطلق عليه الانحرافات العشوائية والتي ترجع إلى الأخطاء العشوائية. ويمكن إجراء اختبار مستوى دلالة معامل الانحدار (م) على النحو التالى:

$$t = \frac{r}{q_m} \text{ حيث } q_m \text{ هى الخطأ المعيارى لمعامل الانحدار (م)}$$

$$\frac{\sqrt{\frac{\text{مربع مجموع انحرافات الانحدار}}{\text{عدد المفردات} - 2}}}{\text{مجموع مربع انحرافات المتغير المستقل عن متوسطه الحسابى}} = \text{قيمة } q_m$$

$$= \frac{\sqrt{\frac{\text{مجموع } (ص - \hat{ص})^2}{n - 2}}}{\text{مجموع } (س - \bar{س})^2}$$

وتسمى القيمة $\frac{\text{مجموع } (ص - \hat{ص})^2}{n - 2}$ تباين الانحدار الذى يحسب من قيم (ص) المتوقعة التى نحصل عليها بتطبيق معادلة خط الانحدار السابقة (راجع صفحة)، ثم نحسب انحرافات الانحدار كما يلى:

ص (ص)	ص (المتوقعة)	ص (المشاهدة)
٥,١٠	٤,٣٠	٩,٤٠
٥,٨٠	٥,٠٥	١٠,٨٥
٥,٧٢	٤,٩٧	١٠,٦٩
٥,٨٥	٥,١٢	١٠,٩٧
٦,٣٢	٥,٦٢	١١,٩٤
٩,٥٦	٩,١٣	١٨,٦٩
٨,٨٦	٨,٣٧	١٧,٢٣
٩,٤٣	٨,٩٨	١٨,٤١
١٠,٢٦	٩,٨٨	٢٠,١٤
١٠,٥٢	١٠,١٦	٢٠,٦٨
<hr/> ٧٧,٤٢		

$$\frac{\chi^2(77,42)}{8} = \text{ق م} = 179,64$$

حيث أن ١٧٩,٦٤ هي قيمة مجد (س - س) من المثال السابق (راجع فصل تحليل الارتباط).

$$,152 = \frac{27,37}{179,64} =$$

$$3,42 = \frac{,52}{,152} = \frac{2}{\text{ق م}} = \text{إذن ت}$$

وحيث أن قيمة (ت) النظرية في جداول توزيع ت بمستوى دلالة ٠,١.

بدرجات حرية (ن - ٢) = (١٠ - ٢) = ٨ هي ٣,٣٦، وحيث أن قيمة ت المحسوبة ٣,٤٢ أكبر من قيمة (ت) النظرية لذلك نرفض الفرض القائل بأنه لا توجد أية علاقة بين المتغيرين (س، ص). أو بعبارة أخرى نقبل الفرض البديل وهو أن هناك علاقة بين المتغيرين (س، ص) في مستوى دلالة أو معنوى ٠,٠١، أى أن هناك احتمال قدره ٩٩% أن لا تكون العلاقة بين المتغيرين (س، ص) قد حدثت بفعل الصدفة أو العشوائية.

وبعد أن أوضحنا كيفية حساب معادلة خط الانحدار بطريقة المربعات الصغرى، وتمثيلها بيانياً وذلك لتحديد نوع العلاقة بين المتغيرين (س، ص)، يبقى أن نحدد قوة هذه العلاقة. وتقاس قوة لعلاقة بين المتغير التابع والمتغير المستقل بواسطة ما يسمى بمقياس معامل التحديد Determination coefficient، ويرمز له بالرمز (ر^٢)، وهو عبارة عن مربع معامل الارتباط فإذا كان معامل الارتباط بين (س، ص) هو ٧٥، مثلاً فإن معامل التحديد يساوى ٢(٧٥) = ٥٦، والفكرة وراء حساب معامل التحديد هي قياس مدى الاختلاف في قيم ص التي ترجع إلى الاختلاف في قيم (س)، وبناء عليه إذا كانت العلاقة بين س، ص قوية فإن ذلك يعنى ارتفاع قيمة معامل التحديد. ونحسب قيمة معامل التحديد ر^٢ بالمعادلة الآتية:

$$r^2 = \frac{\text{م [محد (س - \bar{س}) (ص - \bar{ص})]}}{\text{محد (ص - \bar{ص})}^2}$$

وإذا رجعنا إلى المثال السابق الخاص بالعلاقة بين مساحة حوض النهر وطوله وحسبنا معامل التحديد لهذين المتغيرين لكان هو:

$$r^2 = \frac{(٩٩,١٣,٥٢)}{٦٧,٢٣} = \frac{٥١,٥٥}{٦٧,٢٣} = ٠,٧٦٧$$

أى أن ٠,٧٧ (٧٧ %) من الاختلاف فى قيم (ص) يمكن تفسيرها
بالاختلافات فى قيم «س»، وأن ٠,٣٣ (٣٣ %) من هذه الاختلافات ترجع إلى
عامل الصدفة والأخطاء العشوائية.

الفصل الثالث عشر
تحليل لسلاسل الزمنية
Time Series

الفصل الثالث عشر

تحليل السلاسل الزمنية

Time Series

هناك كثير من الظواهر التي تتغير نحو الزيادة أو النقص بمرور الزمن، وإذا ما تتبعنا مشاهدتها فإننا نحصل على سلسلة Series من هذه المشاهدات. غير أننا غالباً ما نجد أن هذه المشاهدات لا تتزايد أو تتناقص باستمرار، ولكنها تتذبذب بين الزيادة إلى النقص أو العكس على حسب الفترات الزمنية المحددة للدراسة. وتسمى مثل هذه السلسلة من المشاهدات بالسلسلة الزمنية. ومن ثم فإن السلسلة الزمنية هي مجموعة من المشاهدات لظاهرة ما أخذت في فترات زمنية محددة، عادة على فترات أو أبعاد زمنية متساوية (منتظمة). وهى بذلك تحتوى على متغيرين، أحدهما الزمن (المتغير المستقل: z)، والآخر هو قيمة الظاهرة (المتغير التابع: v)، وبعبارة أخرى تعرف السلسلة الزمنية رياضياً بالقيم $v_1, v_2, v_3, \dots, v_n$... ص ن والتي يأخذها المتغير v عند الزمن $z_1, z_2, z_3, \dots, z_n$ ، أى أن v دالة z ، ويرمز لذلك بالرمز $v = f(z)$.

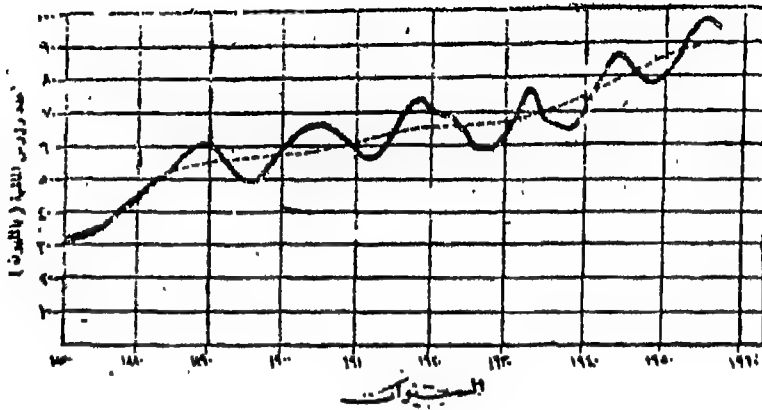
والغرض من دراسة السلسلة الزمنية هو التعرف على التغيرات الكمية التي تطرأ على الظاهرة عبر الزمن لمعرفة أسبابها ونتائجها وما يمكن أن يكون هناك من علاقة بينها وبين غيرها من الظواهر، وكذلك التنبؤ الاحصائي بقيمتها غير المشاهدة أو المرصودة، ومما لذلك من أهمية فى اتخاذ القرارات التي تتعلق بتخطيط المستقبل. ومن أمثلة السلاسل الزمنية الاستهلاك المحلى فى السنة من محصول القطن فى مصر على مدار عدد من السنوات، الإنتاج الكلى فى السنة من خامات الصلب فى

المملكة المتحدة على فترة زمنية تتكون من عدد من السنوات، تسمية الأمطار الساقطة في السنة على أحد المراصد الجوية في مدينة ما لعدة سنوات، عدد السفن التي تعبر قناة السويس على مدار عدد من السنين، عدد سكان مصر في التعدادات المتتالية إلخ.

وستعرض في هذا الفصل لدراسة تطور الظواهر الجغرافية على مدار فترات زمنية متتالية بغرض وصف هذا التطور، والتعرف على طرق قياس التغيرات المختلفة التي تطرأ على هذا التطور بفعل العوامل المتنوعة والمؤثرات العديدة مما يفيد في عملية التنبؤ بما يمكن أن تكون عليه بيانات الظواهر المختلفة في فترات زمنية مضت أولم تأت بعد.

التمثيل البياني للسلاسل الزمنية:

ولغرض قياس التسلسل الزمني يمكن تمثيل السلسلة الزمنية المتضمنة الظاهرة (المتغير ص) مقابل عنصر الزمن (ز) بيانياً كما فعلنا ذلك من قبل (راجع الفصل الثالث). وتعتبر الرسوم البيانية الخطية Line Graph أهم أنواع الرسوم البيانية التي تستعمل لبيان العلاقة بين الظاهرة (المتغير) موضع التحليل ومدى ارتباطها بعنصر الزمن، بحيث توضح أية نقطة على الرسم البياني مقدار هذه الظاهرة خلال فترة زمنية معينة. وهناك نوعان من الرسوم البيانية الخطية يختص النوع الأول منها بالتغيرات الموجبة (نمو مضطرد Growth) والتغيرات السالبة (تناقص أو اضمحلال Decline) لظاهرة معينة، بينما يوضح النوع الثاني التغيرات النسبية Proportional Changes للظاهرة. والنوع الأول يعتمد في تمثيله على التغيرات الحسابية (العددية) Numerical Changes المبنية على الأرقام الفعلية للظاهرة قيد التحليل. وفي هذا النوع من الرسوم يسمى الخط البياني الذي يصل النقاط التي تمثل قيم الظاهرة «بالمنحنى التاريخي» Historigram الذي يستخدم لدراسة الاتجاه العام للظاهرة قيد الدراسة (شكل رقم ١٣-١).



شكل رقم (١٣-١)

المتحني التاريخي لسلسلة أعداد والمأشبة

في الولايات المتحدة الأمريكية (١٨٧٠ - ١٩٦٠) ويعاب على هذا النوع من الرسوم البيانية أن دراسة وتحليل الخط البياني أو المتحني التاريخي تقودنا إلى نتائج مضللة أحياناً، خاصة إذا كان الغرض من التحليل هو مقارنة هذه النتائج العددية مع غيرها من القيم العددية للظواهر الأخرى.

ولبيان وتمثيل التغيرات النسبية في السلسلة الزمنية تستخدم طريقة الأرقام القياسية Index Numbers التي تعتمد على اختيار سنة الأساس Base Year ويعطى لها الرقم ١,٠ ثم تحول بقية القيم في السلسلة إلى نسب مئوية منسوبة إلى سنة الأساس المختارة. فمثلاً إذا أردنا تحويل القيم العددية في الجدول التالي (جدول رقم ١٣-١) إلى أرقام قياسية تتبع الخطوات التالية:

جدول رقم (١٣-١)
تطور أعداد رؤوس الماشية والأغنام في اسكتلندا
(بالألف رأس)

السنة	عدد الماشية	عدد الأغنام	السنة	عدد الماشية	عدد الأغنام
١٩٤٦	١٤٧٢	٦٢٥٤	١٩٥٨	١٨٢٠	٧٩٢٩
١٩٤٨	١٤٩٩	٦٧٣١	١٩٦٠	٢٠٠٣	٨٤٠٧
١٩٥٠	١٦١٦	٧٣٣٧	١٩٦٢	٢٠١٧	٨٦٣٩
١٩٥٢	١٥٧٦	٧٢٧٣	١٩٦٤	١٩٩٠	٨٥٣١
١٩٥٤	١٧١٠	٧٤٢٩	١٩٦٦	٢٠٩١	٨٣٧٧
١٩٥٦	١٧٣٦	٧٥٢٥			

١- نختار سنة ولتكن سنة ١٩٤٦ ونعطيها رقماً قياسياً مقداره ١٠٠ (ويمكن اختيار أى سنة من سنوات السلسلة لتكون سنة الأساس، وقد يؤخذ متوسط عدد السنين كسنة الأساس).

٢- نحول أعداد كل من الماشية والأغنام منسوب إلى سنة الأساس. فمثلاً عدد الأغنام في عام ١٩٤٨ يساوى ٦٧٣١٠٠٠، وعددها في عام ١٩٤٦ هو ٦٣٥٤٠٠٠، ونسبة عدد الأغنام في عام ١٩٤٨ إلى عددها ١٩٤٦ هو:

$$7.98 - 1300 \times \frac{6731000}{6354000}$$

٣- بنفس الطريقة السابقة يمكن الحصول على الأرقام القياسية لأعداد كل من الماشية والأغنام باعتبار سنة ١٩٤٦ هي سنة الأساس لكل من الماشية والأغنام كما في الجدول رقم (١٣-٢).

جدول رقم (١٣-٢)

الأرقام القياسية لتطور أعداد الماشية والأغنام فى اسكتلندا

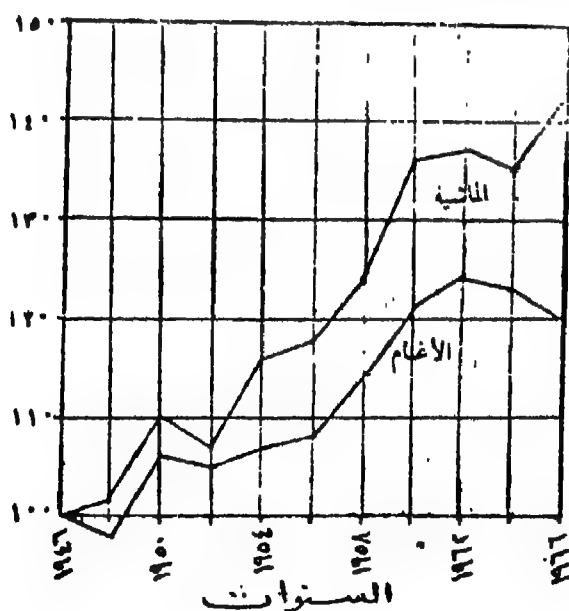
السنة	الأرقام القياسية		السنة	الأرقام القياسية	
	عدد الماشية	عدد الأغنام		عدد الماشية	عدد الأغنام
١٩٤٦	١٠٠	١٠٠	١٩٥٨	١٢٤	١١٤
١٩٤٨	١٠٢	٩٨	١٩٦٠	١٣٦	٢١
١٩٥٠	١١٠	١٠٦	١٩٦٢	١٣٧	١٢٤
١٩٥٢	١٠٧	١٠٥	١٩٦٤	١٣٥	١٢٣
١٩٥٤	١١٦	١٠٧	١٩٦٦	١٤٢	١٢٠
١٩٥٦	١١٨	١٠٨			

٤- تقوم بتمثيل الأرقام القياسية بيانياً بطريقة الخطوط البيانية كما فى الشكل رقم (١٦-٢) والذي يمثل الأرقام القياسية لتطور أعداد الماشية والأغنام من عام ١٩٤٦ حتى ١٩٦٦.

وتتميز طريقة الأرقام القياسية بسهولة قراءة وملاحظة التغير النسبى للظاهرة خلال فترة الدراسة. فمثلاً الرقم القياسى ١٣٦ يعطينا دلالة على نمو وتطور الظاهرة مقدارة ٣٦% عن سنة الأساس، والرقم القياسى ٩٨ يعنى نقص أو اضمحلال الظاهرة موضع الدراسة بما يساوى ٢% عن سنة الأساس.

ومن مميزات هذه الطريقة أيضاً أنها تلغى القيم الأصلية وتحولها إلى نسب مئوية من سنة الأساس مما يسهل عملية المقارنة ودراسة أوجه الشبه والاختلاف بين جميع القيم للظواهر موضع التحليل. فمثلاً من الشكل رقم (١٣-٢) يمكن أن نوضح أن سرعة تزايد أعداد الماشية خلال فترة السلسلة الزمنية أكبر من سرعة تزايد أعداد الأغنام. وذلك على عكس ما يظهر عند دراسة الرسم البيانى الحسابى أو العددي (شكل رقم ١٣-١) مما يؤكد أن الرسوم البيانية التى توضح التغيرات العددية للقيم الأصلية للظاهرة قد تكون مضللة أحياناً. أما أهم عيوب طريقة الأرقام القياسية كوسيلة لبيان تطور الظاهرة أنها لا توضح إلا التغير النسبى بين سنوات السلطة

الزمنية وسنة الأساس فقط. وعلى هذا فإن التغير بين سنوات السلسلة نضمها لا يمكن معرفته إلا إذا كان منسوباً إلى سنة الأساس. فمثلاً فمن الشكل رقم (١٦-٢) نجد أن الأرقام القياسية للماشية والأغنام قد ارتفع بين سنة ١٩٥٦ وسنة ١٩٥٨ بمقدار ٦ (من ١١٨ - ١٢٤ للماشية، ومن ١٠٨ - ١١٤ للأغنام) غير أن انحدار الخط البياني على الرسم للظاهرتين متشابه بين هاتين السنتين. والحقيقة أن هناك اختلافاً واضحاً بين هاتين القيمتين لكل من الظاهرتين، حيث أن ٦ أرقام قياسية من أصل ١١٨ لاتمثل إلا زيادة قدرها ٥٪ في عدد الماشية، في حين أن ٦ أرقام قياسية من أصل ١٠٨ تمثل زيادة قدرها ٦٪ تقريباً في عدد الأغنام. وجدير بالذكر أن اختيار معدل عدد من السنوات (خمس سنوات متتالية مثلاً) كسنة أساس يكون أفضل من اختيار سنة بذاتها لتكون سنة الأساس وذلك لتجنب الذبذبات Fluctuations التي قد تطرأ على تطور ونمو الظاهرة بسبب العوامل المنحكمة والمؤثرة في هذا التطور والنمو.



شكل رقم (١٣-٢)

الأرقام القياسية لتطور أعداد الماشية والأغنام في اسكتلندا

وهناك طريقة أخرى لبيان التغير النسبي قد يطرأ على ظاهرة معينة خلال فترة زمنية محددة هي طريقة الرسم البياني اللوغاريتمي والنصف لوغاريتمي الذي قمنا بشرح فكرة رسمه وإنشائه في الفصل الثالث من هذا الكتاب. ويسود استخدام طريقة الرسم البياني النصف لوغاريتمي في تمثيل معدلات التغير أو النمو للظاهرة والتي تتغير تغيراً زمنياً مثل ظاهرة نمو السكان. فإذا أريد تمثيل التغير الذي يطرأ على مثل هذه الظاهرة يجب استخدام الرسم البياني النصف لوغاريتمي والذي يمثل فيه المحور الأفقي عامل الزمن بالطريقة الحسابية، بينما يمثل المحور الرأسي مقدار تزايد الظاهرة لوغاريتمياً. فمثلاً أهم ما يلاحظ على طريقة إنشاء الشكل رقم (١٣-٣) الذي يوضح النمو السكاني لمدينتين أ، ب لسنوات التعداد من ١٨٠١ إلى ١٩٦١ (جدول رقم ١٣-٣) أن تقسيم المحور الأفقي يتفق مع عدد سنوات التعدادات بينما نجد على المحور الرأسي تقسيماً لوغاريتمياً أي أنه مقسم إلى ثلاث من الدورات اللوغاريتمية تبدأ من ١٠٠٠ إلى ١٠٠٠٠، والثانية من ١٠٠٠٠ إلى ١٠٠٠٠٠، والثالثة من ١٠٠٠٠٠ إلى ١٠٠٠٠٠٠ ولكنها لم تستكمل وانتهت عند رقم ٣٠٠٠٠٠، وينظرة أخرى إلى الشكل نجد على جانبيه اليمين دورتين لوغاريتميتين الأولى تبدأ من ١٠٠ إلى ١٠٠٠ والثانية تبدأ من ١٠٠٠ إلى ١٠٠٠٠. وأهم مميزات الرسم البياني النصف لوغاريتمي السابق أن المقياس على محوريه الرأسيين (الأيسر والأيمن) يساعد على سهولة المقارنة بين ظاهرتين إحدهما مقدارها كبير جداً والأخرى مقدارها صغير جداً. ففي الشكل نجد خطين بيانيين نصف لوغاريتميين يوضحان نمو السكان وتطور عددهم في مدينتين إحدهما ازداد عدد سكانها من ٢٨٨٠١ نسمة إلى ٣١١٨٩٩ نسمة، بينما الأخرى ازداد عدد سكانها من ٨٦٨ نسمة إلى ٥١٨٥ نسمة. وهذا النمط من التطور لا يمكن بيانه بغير الخط البياني اللوغاريتمي أو النصف لوغاريتمي. كما يفيدنا هذا النوع من الرسوم البيانية في معرفة الفترة الزمنية التي تتضاعف فيها قيمة الظاهرة، وذلك بسبب ما تتميز به الأشكال اللوغاريتمية من أن المسافة على المحور الرأسي بين الرقمين ١٠٠٠، ٢٠٠٠ تساوي المسافة التي تقع بين الرقمين ٣٠٠٠، ٦٠٠٠، وتساوي أيضاً المسافة بين الرقمين ٤٠٠٠٠، ٨٠٠٠٠ وكذلك بين الرقمين ٣٠٠٠٠، ٦٠٠٠٠. وبعبارة أخرى أن المسافات على المحور الرأسي

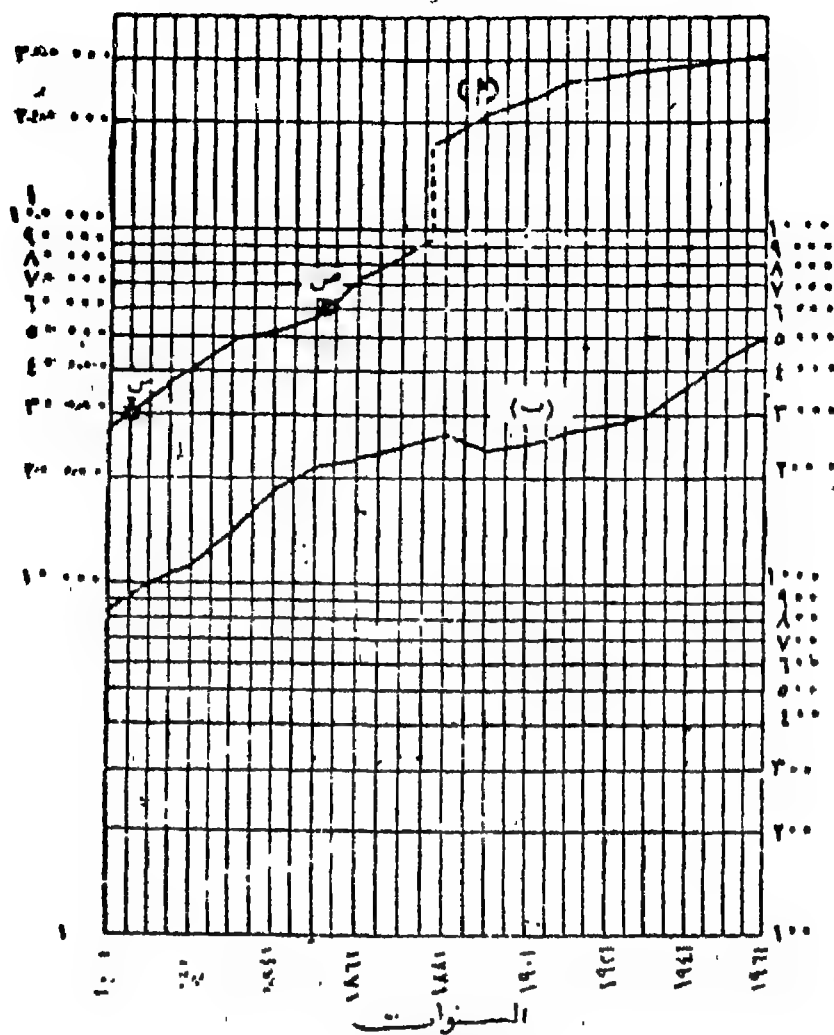
الذى يوضح المقياس اللوغاريتمى تكون متساوية بين أى رقمين النسبة بينها كنسبة ١ : ٢. فمثلاً إذا أردنا معرفة الفترة الزمنية التى يتضاعف فيها السكان فى المدينة «أ» فى الشكل رقم (١٣-٣)، فإننا نأخذ نقطتين على الرسم احدهما تمثل عدد سكان المدينة فى فترة ما، ثم نحدد النقطة الأخرى على الخط البيانى نفسه عندما يتضاعف عدد السكان. فنقطة س تمثل عدد سكان المدينة الذى يبلغ ٣٠٠٠٠ نسمة، ونقطة ص تمثل عدد سكان المدينة حينما بلغ ٦٠٠٠٠ نسمة، فإذا أسقطنا عمودين من هاتين النقطتين على المحور الأفقى نجد أن المسافة بين مسقطيهما تقع بين عامى ١٨٠٦، ١٨٥٣. وبذلك تستنتج أن عدد السكان فى المدينة «أ» قد تضاعف فى النصف الأول من القرن التاسع عشر خلال فترة ٤٧ سنة تقريباً.

جدول رقم (٦-٣)

تعداد السكان لمدينتين أ، ب

١٨٠١ - ١٩٦١

السنة	تعداد السكان		السنة	تعداد السكان	
	المدينة أ	المدينة ب		المدينة أ	المدينة ب
١٨٠١	٢٨٨٠١	١٨٦٨	١٩٠١	٢٣٩٧٤٣	٢٤٩٣
١٨٢١	٤٠١٩٠	١١٣٨	١٩٢١	٢٦٢٦٢٤	٢٨٧٧
١٨٤١	٥٢١٦٤	١٨٣٥	١٩٤١	٢٧٦١٨٩	٣٠٦٤
١٨٦١	٧٤٦٩٣	٢٢٨٣	١٩٦١	٣١١٨٩٩	٥٠٨٥
١٨٨١	٨٦٥٧٥	٢٦٣٨			



شكل رقم (١٣-٣)

تمثيل النمو السكاني لمدينتين أ، ب على الرسم البياني النصف لوغاريتمي

التغيرات (التحركات) المميزة في السلاسل الزمنية:

يمكننا الآن القول - بعد أن عرفنا مميزات الرسم البياني للسلسلة الزمنية أن تطور أو نمو الظاهرة الجغرافية قيد التمثيل هو بمثابة نقطة تتحرك مع مرور الزمن تحت تأثير قوى معينة (طبيعية، بشرية، اقتصادية .. الخ) وذلك فيما يشبه لتحرك

المادى للذرة تحت تأثير قوى مادية مختلفة. ولكن من دراسة وملاحظة كثير من السلاسل الزمنية يمكن اكتشاف وجود تغيرات (تحركات) Changes أو اختلافات مميزة ودرجات مختلفة.. وجدير بالذكر أن تحليل مثل هذه لاتغيرات له أهمية كبرى فى التنبؤ بالتغيرات المستقبلية الظاهرة قيد الدراسة. وبناء على ذلك يمكن تصنيف التغيرات فى السلاسل الزمنية إلى أربعة أنماط تسمى غالباً «مكونات السلسلة الزمنية» نوجزها فيما يلى:

١- التغيرات طويلة المدى (الاتجاه العام) The Trend. تشير هذه التغيرات إلى الاتجاه العام (الحركة العامة) الذى يكون عليه الشكل البياني للسلسلة الزمنية على مدار فترة طويلة من الزمن. ويرمز للاتجاه العام على الرسم البياني للسلسلة الزمنية بمنحنى أو خط الاتجاه العام الذى يمثل التدرج والانتظام والاستمرار فى السلسلة الزمنية.

٢- التغيرات الدورية Periodical (Cyclical) Variations. وتشير هذه التغيرات إلى الذبذبات طويلة المدى حول خط الاتجاه العام. وتسمى هذه التغيرات أحياناً «دورات» التى قد تتبع أولاً تتبع نفس النمط بعد كل فترة زمنية متساوية. وفى مجال الجغرافية الاقتصادية تعتبر التغيرات دورية إذا تكررت بعد فترات زمنية تزيد عن سنة.

٣- التغيرات الموسمية Seasonal Variation تمثل هذه التغيرات النمط المتماثل، أو المنتظم لحركة السلسلة الزمنية فى الأشهر المتقابلة خلال السنوات المتتالية، وتمثل هذه التغيرات فى لحادثات التى تقع شهرياً أو ربع سنوياً أو سنوياً مثل زيادة مبيعات المحلات التجارية فى الفترة السابقة للأعياد أو موسم الحج، زيادة كمية الأمطار خلال فصل الشتاء عنه أثناء فصل الصيف فى الأقاليم شبه الجافة فى العروض المعتدلة فى غرب القارات. وعلى الرغم من أن التغيرات الموسمية تشير إلى الدورة السنوية، فإنها يمكن أن تمتد لتشمل الدورية لأية فترة من الزمن مثل ساعة، يوم، أسبوع ... الخ اعتماداً على نوع البيانات المتاحة. وأهم أسباب التغيرات الموسمية فى السلسلة الزمنية هى العادات والتقاليد الاجتماعية والعوامل المناخية.

٤ - التغيرات العشوائية Random Variations. وتشير هذه التغيرات إلى الحركة غير المنتظمة في السلسلة الزمنية، بمعنى أنه ليس لها نمط معين أو قاعدة ثابتة، أى أنها تنجم فى الغالب عن عوامل عارضة أو فجائية لا يمكن التنبؤ بمواعيد حدوثها بدقة مثل الفيضانات والزلازل ونتائج الانتخابات ... وغيرها. وعلى الرغم أنه من المعتاد افتراض أن مثل هذه الأحداث تنتج تغيرات تستمر لفترة قصيرة من الزمن فمن المعقول أن تكون على درجة من الكثافة نتيجة لوجود دورات جديدة أو غيرها من التغيرات ولا بد من استبعادها عند قياس الاتجاه العام أو عند قياس التغيرات الموسمية للظاهرة موضع التحليل.

وعموماً فإن الظاهرة الواحدة قد تتأثر بجميع هذه التغيرات أو بعضها الأمر الذى ينجم عنه تشكلات صاعدة أو هابطة فى خط سير الظاهرة. ولهذا السبب يهتم الجغرافيون بدراسة مثل هذه التغيرات ومعرفة أسبابها لتفادى آثارها الضارة فى المستقبل.

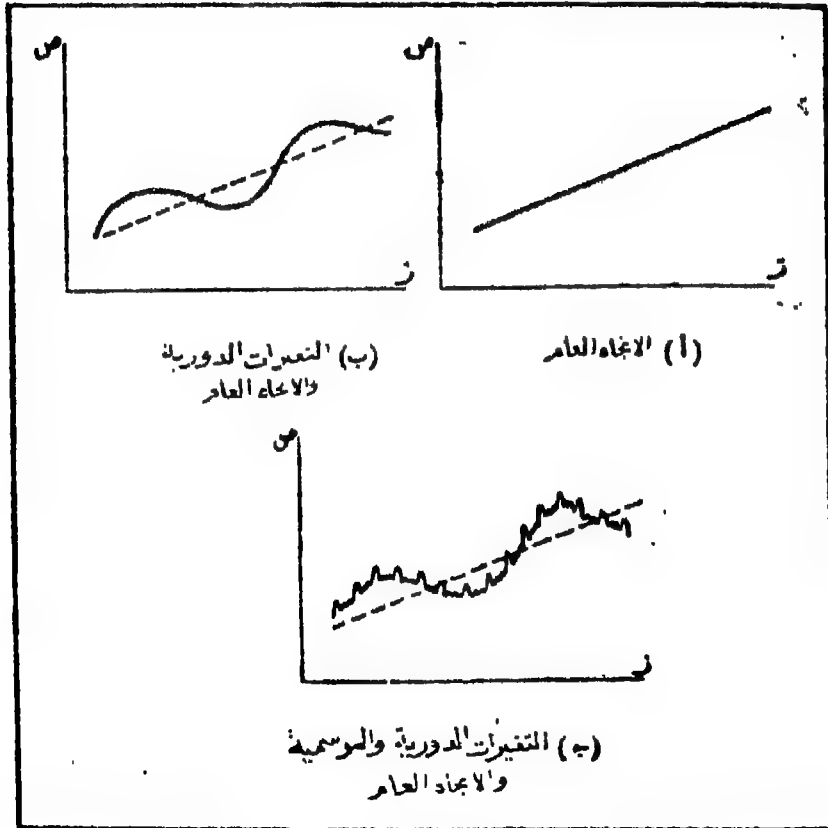
تحليل السلسلة الزمنية:

يهدف تحليل السلسلة الزمنية إلى إيجاد طريقة مناسبة لقياس التغيرات وبالتالي دراسة علاقاتها بالظروف المختلفة. ويتضمن تحليل السلاسل الزمنية وصف مكونات التغيرات الموجودة داخل السلسلة، أو بعبارة أخرى إيجاد مركبات القوى التى تؤثر فى خط سير الظاهرة خلال الفترة الزمنية موضع التحليل. ولتوضيح الطرائق التى تستخدم فى هذا الوصف نشير إلى الشكل رقم (١٣-٤) والذى يعرف بالسلسلة الزمنية المثالية التى تتخذ فيها المركبات (القوى) أشكالاً بيانية مختلفة.

ويمكن أن يعطينا عرض الأشكال البيانية السابقة (شكل رقم ١٣-٤ أ؛ ب، ج) أسلوباً خاصاً لتحليل السلاسل الزمنية. فتفترض أن المتغير (ص) الذى يعبر عن السلسلة الزمنية هو حاصل ضرب المتغيرات ن، د، و، ش والتى ينتج عنها الاتجاه العام (ت) والتغيرات الدورية (د) والتغيرات الموسمية (و) والتغيرات العشوائية (ش). أى أن قيمة الظاهرة تساوى حاصل ضرب مركباتها الأساسية. وباستخدام الرموز نجد أن:

$$ص = ت \times د \times و \times س = ت \times د \times و \times س$$

ومن هذا المنطلق فإن تحليل السلاسل الزمنية يتضمن فحص المتغيرات ت، د، و، ش السابقة والتي يشار إليها بمفكوك السلسلة الزمنية أو المكونات الأساسية لمتغيراتها. وتجدر الإشارة إلى أن بعض الاحصائيين يفضلون اعتبار قيمة الظاهرة مساوية لمجموع قيم مركباتها، أو بعبارة أخرى $ص = ت + د + و + س$. وكل من الطريقتين السابقتين لتحديد قيمة الظاهرة يمكن استخدامها في تحليل السلسلة الزمنية. ولكي نعين تأثير كل قوة من قوى التأثير (قوة الاتجاه العام، قوة التغيرات الدورية، الموسمية العشوائية) على السلسلة الزمنية يجب عزل كل قوة من هذه القوى عن تأثير القوى الأخرى.



شكل رقم (١٣-٤) مكونات السلسلة الزمنية المثالية

أولاً: تقدير الاتجاه العام:

يتناول الجغرافيون دراسة الظواهر التي تتزايد بطبيعتها خلال فترة زمنية مثل كميات الإنتاج للمحاصيل أو السلع الصناعية والمعدنية، وعدد السكان وحجم الاستهلاك، أو الظواهر التي تتناقص على مدار الزمن مثل معدل الوفيات واستهلاك السلع. ويكون أحسن تمثيل بياني للاتجاه العام للسلاسل الزمنية لمثل هذه الظواهر أن نمهد لها بطريقة مناسبة، خطأً مستقيماً أو منحني أملسا يمثل سير الظاهرة لو لم تؤثر عليها عوامل أخرى تؤدي إلى تفاوت في مقدار الظاهرة. وعندما يتم رسم خط - أو تمهيد منحني - الاتجاه العام فإننا نجد أن كل فترة زمنية يناظرها قيمتان للظاهرة إحداهما هي القيمة المشاهدة الحقيقية والأخرى هي القيمة الاتجاهية، وهي القيمة التي يحددها الخط أو المنحني الممهد، وهي بطبيعة الحال تختلف عن القيمة المشاهدة. وتجدر الإشارة إلى أن متوسط القيم الاتجاهية يساوي متوسط القيم المشاهدة. ويمكن معرفة وتقدير الاتجاه العام Trend للظاهرة بعدة طرق هي: طريقة الخط (منحني) الاتجاه العام وتمهيده باليد، طريقة أنصاف المتوسطات (أشباه المتوسطات)، طريقة المتوسطات المتحركة، طريقة المربعات الصغرى. وسنلقى الضوء فيما يلي على كل طريقة على حدة لإبراز أهم مميزاتها وعيوبها عند اتخاذها كأساس لتحليل السلسلة الزمنية.

١- طريقة التمهيد باليد: تعتمد هذه الطريقة على توفيق أو تمهيد خط الاتجاه العام باليد بحيث يمر بأكبر عدد ممكن من النقاط - التي تمثل قيم الظاهرة في كل وحدة زمنية على الرسم البياني - أو يمر بين النقاط باتزان وذلك بعد رسم المنحني التاريخي للسلسلة عن طريق توصيل النقاط بخطوط منكسرة. ويمكن أن نستنتج من الرسم البياني بصورة عامة شكل الاتجاه العام، بمعنى هل هو خط مستقيم فيكون له معادلة من الدرجة الأولى، أو منحني من نوع معين قد تكون معادلته من الثانية أو الثالثة أو من درجة أعلى من ذلك.

ويمثل الشكل رقم (١٣-٥) منحنيًا تاريخيًا وخطاً ممهداً باليد للسلسلة الزمنية التي توضح تطور إجمالي الصادرات لجمهورية مصر في الفترة ١٩٦٠ - ١٩٧٢ (جدول رقم ١٣-٤).

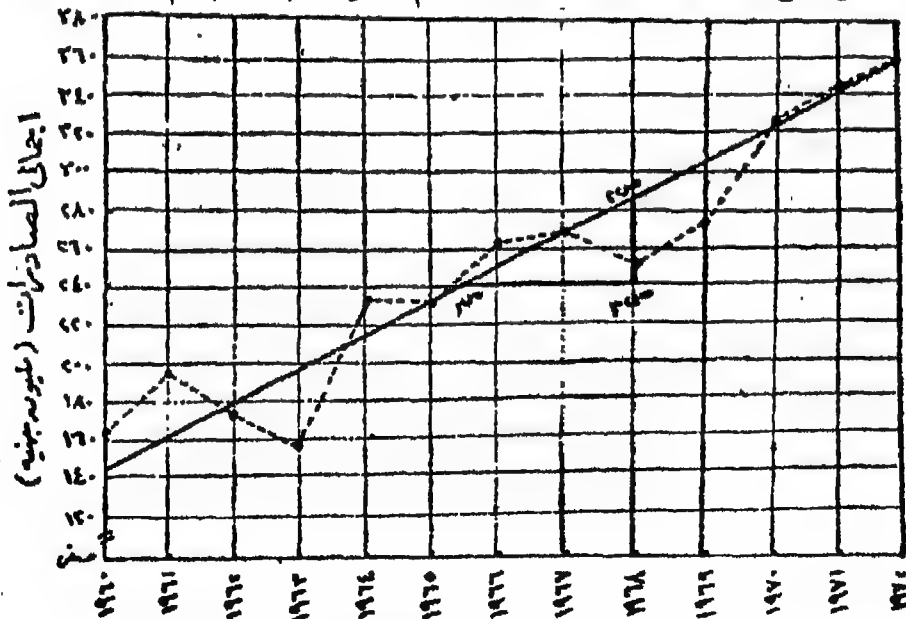
جدول رقم (١٣-٤)

تطور إجمالي الصادرات لجمهورية مصر (بالألف جنيه مصرى)

فى الفترة ١٩٦٠ - ١٩٧٢

الصادرات	السنة	الصادرات	السنة	الصادرات	السنة
٣٢٣٩٢٩	١٩٧٠	٢٣٤٣٧٧	١٩٦٥	١٦٠٤٥٥	١٩٦٠
٣٤٣١٧٧	١٩٧١	٢٦٣١٣٢	١٩٦٦	١٩٧٧٥٧	١٩٦١
٣٥٨٧٧٥	١٩٧٢	٢٦٣١٣٥	١٩٦٧	١٦٨٩٢٦	١٩٦٢
		٢٤٦١٣٧	١٩٦٨	١٥٨٥١٠	١٩٦٣
		٢٧٠٢٩٥	١٩٦٩	٢٣٦٧٩٨	١٩٦٤

ومن الخط الممهد يمكن حساب مقدار التغير الذى يطرأ على الظاهرة فى كل وحدة زمنية فى المتوسط - وهو ما يطلق عليه «معدل التغير» - وذلك بأن نحدد نقطتين على الخط الممهد (خط الاتجاه العام) مثل ص ١، ص ٢ ثم من النقطة ص ٢



شكل رقم (١٦-٥) الاتجاه العام لتطور إجمالي الصادرات

لجمهورية مصر فى الفترة ١٩٦٠ - ١٩٨٠

نرسم عموداً يوازي المحور الرأسى، ومن النقطة ص_١ نرسم خطاً يوازي المحور الأفقى بحيث يلتقى الخطان عند النقطة ص_٣ فيكون معدل التغير هو ميل الخط الممهد (أى ظل الزاوية ص_١) وهو يساوى الاحداثى الرأسى ص_٢، ص_٣ بوحدات المحور الرأسى (إجمالى الصادرات بالمليون) مقسوماً على الاحداثى الأفقى ص_١، ص_٣ بوحدات المحور الأفقى (الوحدات الزمنية). ومن الشكل رقم (١٣-٥) يتبين أن معدل إجمالى الصادرات فى المتوسط ١٨,٥ مليون جنيه. وعلى فرض أن الاتجاه العام يأخذ شكل الخط المستقيم الذى معادلته ص = م س ± ح، فإن معدل التغير (م)، وهو ميل الخط، يكون ثابت على جميع أجزاء الخط مهما اختلفت الفترة الزمنية المحسوب عندها المعدل. وبذلك يمكن تحديد المعادلة السالفة بمعلومية معدل التغير (م) والجزء المقطوع من المحور الرأسى (ح) والذى يمثل هنا الحد الأدنى لإجمالى الصادرات سنوياً (يمكن تحديده من الرسم بالمقدار ١٢٥ مليون جنيه). وعلى ذلك فإن معادلة الخط المستقيم لهذه السلسلة تكون:

$$ص = ١٨,٥ س + ١٢٥$$

ويعاب على طريقة تعيين الاتجاه العام التمهيد باليد بأنها ليست دقيقة إذ أن العمل فيها يتم بطريقة تقديرية تختلف من شخص إلى آخر، كما تتوقف على مدى خبرة ومران الباحث فى تمهيد خط مستقيم باليد ليمر بأكبر عدد ممكن من النقاط. وعلى الرغم من ذلك فإن خط الاتجاه العام الممهد باليد يمكن استخدامه كمؤشر للتنبؤ بقيم الظاهرة فى فترة لاحقة لفترة السلسلة الزمنية موضع التحليل، إذ أن أحداثى أية نقطة تقع على خط الاتجاه عبارة عن إجمالى الصادرات عند وسعة زمنية مناظرة. فعلى سبيل المثال يمكن حساب القيمة المقدرة لإجمالى الصادرات لعام ١٩٨٥ كالتالى:

$$\text{القيمة الاتجاهية (ص)} = ١٨,٥ س + ١٢٥$$

قيمة س (فى هذه الحالة على أساس أن نقطة الأصل هى ١٩٦٠)

$$١٩٨٥ - ١٩٦٠ = ٢٥ \text{ بالتعويض فى معادلة خط الاتجاه العام نجد:}$$

$$ص = ١٨,٥ \times ٢٥ + ١٢٥ = ٥٨٧,٥ \text{ مليون جنيه}$$

٢- طريقة أنصاف (أشباه) المتوسطات Semi-Average: وهي طريقة تقريبية أيضاً لتعيين خط الاتجاه العام ولكنها أفضل من التمهيد باليد. ولتوفيق خط مستقيم يوضح الاتجاه العام لسير الظاهرة تقسم البيانات إلى قسمين متساويين - كلما أمكن - ثم نحصل على المتوسط الحسابي لكل قسم على حدة، وهذا يعطينا قيمة نقطتين على خط السلسلة الزمنية (في حالة إذا كانت السلسلة تحتوى على عدد زوجي من السنين فإنه يمكن تقسيمها إلى قسمين متساويين بإهمال السنة التي تقع في منتصف السلسلة). ولرسم خط الاتجاه العام نعين على الرسم النقطتان التي تمثل كل منهما قيمة كل متوسط أمام منتصف الفترة الزمنية لكل مجموعة ثم نصل بين هاتين النقطتين، وبذلك يمكن تحديد القيم الاتجاهية.

مثال (٣):

تبين البيانات في الجدول رقم (١٣-٥) الإنتاج السنوي من الفحم (مليون كيلو جرام) في دولة ما بين سنة ١٩٤٨ وسنة ١٩٥٨ وتحليلها بطريقة أنصاف المتوسطات بهدف تحديد الاتجاه العام للإنتاج خلال هذه الفترة تجرى الخطوات التالية:

جدول رقم (١٣-٥)

الإنتاج السنوي من الفحم (مليون كيلو جرام)

في دولة ما الفترة ١٩٤٨ - ١٩٥٨

السنة	١٩٤٨	١٩٤٩	١٩٥٠	١٩٥١	١٩٥٢	١٩٥٣	١٩٥٤	١٩٥٥	١٩٥٦	١٩٥٧	١٩٥٨
الإنتاج	٥٠,٠	٣٦,٥	٤٣,٠	٤٤,٥	٣٨,٩	٣٨,١	٣٢,٦	٣٨,٧	٤١,٧	٤١,١	٣٣,٨

١- تحتوى هذه السلسلة الزمنية على بيانات ١١ سنة فنقسمها إلى قسمين متساويين كل منهما ٥ سنوات (مع حذف السنة في منتصف السلسلة وهي سنة ١٩٥٣)، فيحتوى القسم الأول على السنوات ١٩٤٨ - ١٩٥٢، والقسم الثانى على السنوات ١٩٥٤ - ١٩٥٨.

٢- مجموع الإنتاج السنوى للقسم الأول يساوى ٢١٢,٩ ، ومجموع الإنتاج السنوى للقسم الثانى يساوى ١٨٧,٩ .

٣- المتوسط الحسابى للقسم الأول $= 212,9 \div 5 = 42,6$ ، والمتوسط الحسابى للقسم الثانى $= 187,9 \div 5 = 37,6$. توقع القيمتان على الرسم البيانى (شكل رقم ١٣-٦) مقابل السنة الوسطى فى القسم الأول وهى سنة ١٩٥٠ ، ومقابل السنة الوسطى فى القسم الثانى وهى سنة ١٩٥٦ (س، س) على الترتيب) ، فإذا وصلنا بينهما فإن الخط الذى يمر بينهما هو خط الاتجاه العام للظاهرة ، والنقط على هذا الخط التى تقابل السنين هى القيم الاتجاهية لهذه السنين .

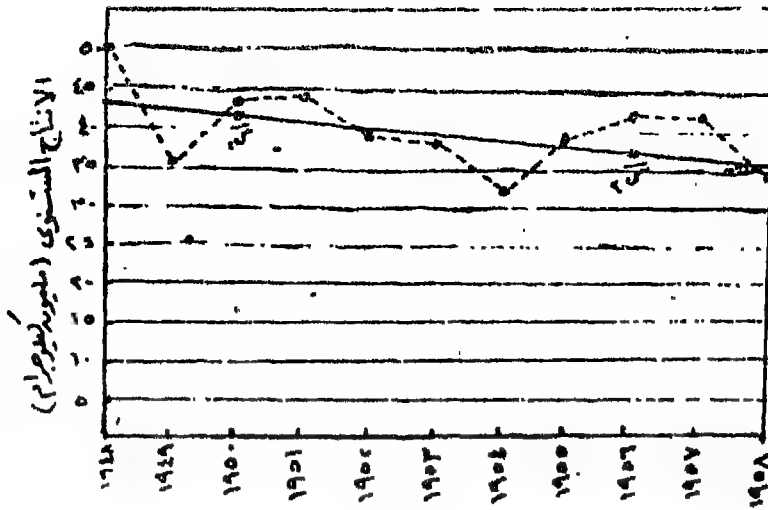
٤- تعد القيمة الاتجاهية لسنة ١٩٥٠ هى ٤٢,٦ مليون جرام ، والقيمة الاتجاهية لسنة ١٩٥٦ هى ٣٧,٦ مليون جرام ، فإذا كانت النتائج التى حصلنا عليها تدل على أنه فى ٦ سنوات (١٩٥٠ - ١٩٥٦) حدث إنخفاض يساوى ٥ مليون كيلو جرام (٤٢,٦ - ٣٧,٦) أى بانخفاض نسوى مقداره $5,0 \div 6 = 0,83$ مليون كيلو جرام / سنة ، وهو ما أطلقنا عليه سلفاً معدل التغير (ميل الخط المستقيم) . وبعبارة أخرى إذا كان خط الاتجاه العام خطاً مستقيماً فإن ميله (م) يساوى الفرق بين المتوسطين الحسابيين لمتتصفى السلسلة الزمنية مقسوماً على الفرق بين زمنيتهما ، أى أن :

$$م = \frac{0,83}{6} = \frac{42,6 - 37,6}{1950 - 1956}$$

٥- إذا كانت معادلة خط الاتجاه العام هى $ص = م س \pm ح$ ، وإذا أخذنا نقطة عند سنة ١٩٥٠ فإن (ح) تساوى المتوسط الحسابى للقسم الأول وهو ٤٢,٦ وتكون معادلة الاتجاه العام هى : $ص = 0,83 س + 42,6$.

٦- لتعيين القيمة الاتجاهية لأية سنة يجب أن تأخذ فى الحسبان السنة التى أخذنا عندها نقطة الأصل فمثلاً سنة ١٩٥١ تقابل $س = 1$ ، فتكون القيمة الاتجاهية لسنة ١٩٥١ هى :

$$ص = 0,83 \times (1 - 1) + 42,6 = 42,6 + 0,83 = 41,77$$



شكل رقم (١٣-٣)

المنحنى التاريخي وخط الاتجاه العام بطريقة أنصاف المتوسطات
للإنتاج السنوي من القمح في دولة ما للفترة ١٩٥٨-٤٨

وإذا أخذنا نقطة الأصل عند سنة ١٩٥٦ فإن (ح) تساوى المتوسط الحسابي
للقسم الثاني (٣٧, ٦) وتكون معادلة الاتجاه العام هي : ص = ٨٣, ٦ + ٣٧, ٦
وتكون القيمة الاتجاهية لسنة ١٩٥٨ (التي تقابل س = ٢-) كالآتي:

$$ص = ٨٣, ٦ \times (-٢) + ٣٧, ٦ = ١, ٦٦ - ٣٧, ٦ = -٣٥, ٩٤$$

٧- من الشكل البياني (شكل رقم ١٣-٦) يتضح أن هناك إجهاداً عاماً هابطاً
Downward Secular Trend ولكنه هبوط بسيط جداً، أى أن الإنتاج السنوي
إنخفض بشكل ضئيل في غضون الفترة الزمنية موضع الدراسة. ومن الواضح
كذلك أن اختلافات القيم الأصلية وانحرافها عن خط الاتجاه العام يمكن
تلخيصها بالإشارات الموجبة والسالبة الآتية: - + + + - - - + + - +
وإذا كان الانحراف عن خط الاتجاه العام شديداً فإنه يجب علينا أن نحدد
طبيعته، بمعنى هل هذا التفاوت والانحراف عن خط الاتجاه العام عشوائي؟ أم
أنه يتبع نظاماً محدداً أو نمطاً معيناً؟

وعلى الرغم من أن هذه الطريقة بسيطة فى تطبيقها، إلا أنها قد تؤدي إلى نتائج غير دقيقة إذا استخدمت بدون تمييز. فكما عرفنا أن حساب ميل خط الاتجاه لعام يتوقف على حساب كل من المتوسطين الحسابين ولا يخفى علينا أن حسابهما يتأثر بالقيم المتطرفة أو الشاذة (الشديدة الارتفاع أو الشديدة الانخفاض أو الاثنين معاً)، فإذا كان أحد القسمين يحتوى على قيم مرتفعة أو منخفضة أكثر من القسم الآخر فإن ذلك يؤثر على قيمتى المتوسطين الحسابين وبالتالي لا يكون خط الاتجاه العام المحسوب مثلاً الحالة بدقة كاملة. ومن عيوب هذه الطريقة أيضاً أنها قابلة للتطبيق فقط فى حالة إذا كان الاتجاه خطأً أو يقرب إلى خطين، على الرغم من أنه يمكن مد صلاحيتها فى الحالات التى يمكن تقسيم البيانات فيها إلى عدد من الأقسام فى كل قسم يكون الاتجاه العام فيه خطياً.

٣- طريقة المتوسطات المتحركة The Moving Averages: عرفنا الغرض الأساسى من تعيين خط الاتجاه العام هو تقدير قيم الاتجاهية لانتقع إلا تحت تأثير عامل واحد فقط هو أثر الاتجاه العام، أى تقليل أثر المتغيرات الدورية، الموسمية والعشوائية. وقد وجدنا أن تعيين الاتجاه بالتمهيد باليد وبطريقة أنصاف المتوسطات ليست بالدقة الكافية التى يعتمد عليها فى التنبؤ. وحفاظاً على الهدف الرئيسى من الاتجاه العام للظاهرة يمكن تعيين الاتجاه العام بأسلوب أسهل باستخدام مايسمى «بالمتوسطات المتحركة»، وهى عبارة عن استخراج متوسط قيم عدد معين (ثلاث أو خمس سنوات مثلاً) من السنوات المتعاقبة أو المتداخلة (فترة فى السلسلة الزمنية)، ثم نثبت قيم هذه المتوسطات أمام السنوات الوسطى لكل فترة فى السلسلة الزمنية. ويمكن تحديد طول الفترة فى السلسلة الزمنية من ملاحظة المنحنى التاريخى وذلك عن طريق معرفة الفرق الزمنى بين كل قيمتين متتاليتين أو كل قاعدتين متتاليتين.

فمثلاً إذا فرضنا أن طول كل فترة فى السلسلة (ن) يساوى ٣ سنوات وأن قيم السلسلة الزمنية هى : ١، ٢، ٣، ٤، ٥، ٦، ٧، ٨، ٩، ١٠، ١١، ١٢، ١٣، ١٤، ١٥، ١٦، ١٧، ١٨، ١٩، ٢٠، ٢١، ٢٢، ٢٣، ٢٤، ٢٥، ٢٦، ٢٧، ٢٨، ٢٩، ٣٠، فإن المتوسطات المتحركة - أو القيم الاتجاهية - من الدرجة ن يمكن أن نحصل عليها بمتابعة من المتوسطات الحسابية:

(أ) القيمة الاتجاهية الأولى = $\frac{١س + ٢س + ٣س}{٣}$ وتمثل قيمة الاتجاهية للسنة ٢س

(ب) القيمة الاتجاهية الثانية = $\frac{٢س + ٣س + ٤س}{٣}$ وتمثل قيمة الاتجاهية للسنة ٣س

(ج) القيمة الاتجاهية الثالثة = $\frac{٣س + ٤س + ٥س}{٣}$ وتمثل قيمة الاتجاهية للسنة ٤س

(د) القيمة الاتجاهية الأولى = $\frac{١س-٢س + ٢س-٣س + ٣س-٤س}{٣}$ وتمثل القيمة الاتجاهية للسنة قبل الأخيرة والتي ترتيبها م-١

وتسمى المجاميع ١س + ٢س + ٣س ، ٢س + ٣س + ٤س ، ٣س + ٤س + ٥س ،
٢س-١س + ٣س-٢س + ٤س-٣س بالمجاميع المتحركة من الدرجة ن.

أما إذا كانت (ن) زوجية، أى ٤ سنوات مثلاً، فإن:

(أ) القيمة الاتجاهية الأولى = $\frac{١س + ٢س + ٣س + ٤س}{٤}$ وتمثل القيمة الاتجاهية

غير المركزية أى لمنتصف السنة الثانية (٢س) ويرمز لها بالرمز ل١

(ب) القيمة الاتجاهية الثانية = $\frac{٢س + ٣س + ٤س + ٥س}{٤}$ وتمثل القيمة الاتجاهية

غير المركزية أى لمنتصف السنة الثالثة (٣س) ويرمز لها بالرمز ل٢.

وحتى نحصل على القيم الاتجاهية لسنوات السلسلة نفسها بدلاً من منتصف سنوات السلسلة فإننا نحسب القيم الاتجاهية على أساس أنها متوسط القيم لكل فترتين متتاليتين. وتكون القيمة الاتجاهية للسنة الثالثة فى السلسلة $\frac{١ل + ٢ل}{٢}$

ويمكن التوصل إلى نفس النتيجة بقسمة المجموع المتحرك لكل فترتين طول كل منهما ٤ سنوات على مجموع سنوات الفترتين وهى ٨ سنوات. وهكذا فإننا نحصل على متوسطات حسابية تتحرك أمام كل سنة من سنوات السلسلة الزمنية

وتمثل قيمها الاتجاهية. ومن المعتاد أن نضع قيمة المتوسطات المتحركة في مكانها الملائم بالنسبة للبيانات الأصلية كما في الشكل رقم (١٣-٧).

السنة	القيمة	القيمة الاتجاهية (متوسط متحرك لثلاث سنوات)	القيمة الإيجارية (متوسط متحرك لخمس سنوات)
١	١ ص	$\frac{١ ص + ٢ ص + ٣ ص}{٣}$	
٢	٢ ص	$\frac{٢ ص + ٣ ص + ٤ ص}{٣}$	$\frac{١ ص + ٢ ص + ٣ ص + ٤ ص + ٥ ص}{٥}$
٣	٣ ص	$\frac{٣ ص + ٤ ص + ٥ ص}{٣}$	$\frac{٢ ص + ٣ ص + ٤ ص + ٥ ص + ٦ ص}{٥}$
٤	٤ ص		
٥	٥ ص		
٦	٦ ص		

شكل رقم (١٣ - ٧)

كيفية حساب القيم الاتجاهية من المتوسطات المتحركة

وباستخدام هذا الأسلوب يمكن استنتاج سلسلة جديدة تستبدل قيم السلسلة الأصلية بمتوسطات لعدد من الفترات الزمنية. فالمتوسطات الحسابية من الدرجة ٥ (أى على أساس ٥ سنوات) تستبدل قيمة كل سنة للسلسلة الأصلية بمتوسط القيم الخاصة لهذه السنة وقيم السنتين السابقتين لها والسنتين اللاحقتين لها، وبالمثل يمكن تكوين سلسلة زمنية جديدة من القيم هي سلسلة المتوسطات المتحركة لأى عدد من السنوات، وهى قيم اتجاهية أقل عدد من القيم الأصلية، ففي حالة ٣ = نجد أنها تبدأ بقيمة أمام السنة الثانية فى السلسلة وتنتهى بقيمة أمام السنة

الأخيرة. وعموماً ما كلما كانت المتوسطات المتحركة محسوبة لعدد أكبر من السنوات كلما حصلنا على قيم أكثر تمهيداً.
مثال:

إذا كانت لدينا السلسلة الزمنية التي تمثل الإنتاج السنوى فى دولة ما من الفحم (بمليون الكيلو جرامات) للسنوات من ١٩٤٨ - ١٩٥٨ التي يوضحها الجدول رقم (١٣-٥)، وعلى فرض أن طول الفترة (ن) هو ٥ سنوات فإن المتوسطات المتحركة للكمية المنتجة للفحم للسلسلة المذكورة يمكن حسابها كما فى الجدول التالى:

جدول رقم (١٣-٦)

حساب القيم الاتجاهية بطريقة المتوسطات المتحركة (على أساس ٥ سنوات) لإنتاج السنوى من الفحم (مليون لكيلو جرام) فى دولة ما فى الفترة من ١٩٤٨ - ١٩٥٨

(١)	(٢)	(٣)	(٤)
السنة	السنة	٥ سنوات مجمرع متحرك	٥ سنوات متوسط متحرك
١٩٤٨	٥٠,٠		
١٩٤٩	٣٦,٥		
١٩٥٠	٤٣,٠	٢١٢,٩	٤٢,٦
١٩٥١	٤٤,٥	٢٠١,٠	٤٠,٢
١٩٥٢	٣٨,٩	١٩٧,١	٣٩,٤
١٩٥٣	٣٨,١	١٩٢,٨	٣٩,٦
١٩٥٤	٣٢,٦	١٩٠,٠	٣٨,٠
١٩٥٥	٣٨,٧	١٩٢,٢	٣٨,٤
١٩٥٦	٤١,٧	١٨٧,٩	٣٧,٦
١٩٥٧	٤١,١		
١٩٥٨	٣٣,٨		

بالنظر إلى الجدول السابق نجد أن المجموع المتحرك الأول (٢١٢,٩) بالعمود

الثالث هو المجموع من العنصر الأول إلى العنصر الخامس في العمود الثاني إلى العنصر السادس وهكذا. ومن الناحية العملية فإنه بعد الحصول على المجموع المتحرك الأول (٢١٢,٩) يمكن الحصول بسهولة على المجموع المتحرك الثاني وذلك بطرح ٥٠,٠ (العنصر الأول في العمود الثاني) وإضافة ٣٨,١ (العنصر السادس في العمود الثاني) فتكون النتيجة ٢٠١,٠. ونفس الطريقة يمكن الحصول على المجاميع المتحركة بسهولة ويسر. وبقسمة كل مجموع متحرك على ٥ ينتج المتوسط المتحرك.

جدول رقم (١٣-٧)

حساب القيم الاتجاهية بطريقة المتوسطات المتحركة (على أساس ٤ سنوات) للإنتاج السنوي من الفحم (مليون كيلو جرام) في دولة ما في الفترة ١٩٤٨ - ١٩٥٨

(١)	(٢)	(٣)	(٤)	(٥)	(٦)	(٧)	(٨)
السنة	البيانات	٤ سنوات مجموع متحرك	٤ سنوات متوسط متحرك غير مركزي	٢ سنة مجموع متحرك للعמוד (٤)	٤ سنوات مجموع متحرك مركزي للعמוד الخامس ÷ ٢	٢ سنة مجموع متحرك للعמוד (٣)	٤ سنوات مجموع متحرك مركزي للعמוד السابع ÷ ٨
١٩٤٨	٥٠,٠						
١٩٤٩	٣٦,٥						
١٩٥٠	٤٣,٠	١٧٤,٠	٤٣,٥	٨٤,٢	٤٢,١	٣٣٦,٩	٤٢,١
١٩٥١	٤٤,٥	١٦٢,٩	٤٠,٧	٨١,٨	٤٠,٩	٣٢٧,٤	٤٠,٩
١٩٥٢	٣٨,٩	١٦٤,٥	٤١,١	٧٩,٦	٣٩,٨	٣١٨,٦	٣٩,٨
١٩٥٣	٣٨,١	١٥٤,١	٣٨,٥	٧٥,٦	٣٧,٨	٣٠٢,٤	٣٨,٨
١٩٥٤	٣٢,٦	١٤٨,٣	٣٧,١	٧٤,٩	٣٧,٥	٢٩٩,٤	٣٧,٥
١٩٥٥	٣٨,٧	١٥١,١	٣٧,٨	٧٦,٣	٣٨,٢	٣٠٥,٢	٣٨,٢
١٩٥٦	٤١,٧	١٥٤,١	٣٨,٥	٧٧,٣	٣٨,٧	٣٠٩,٤	٣٨,٧
١٩٥٧	٤١,١	١٥٥,٣	٣٨,٨				
١٩٥٨	٣٣,٨						

بالرجوع إلى الجدول السابق نلاحظ أننا حصلنا أولاً على المتوسطات المتحركة غير المركزية (على أساس ٤ سنوات)، أى التى تتركز بين السنوات المتتالية. ولحساب المتوسطات المتحركة لسنوات السلسلة نفسها قمنا بحسب مجموع متحرك للمتوسطات المتحركة (على أساس ٢ سنة) كما فى العمود (٥). ثم حصلنا على المتوسطات المتحركة المركزية عن طريق قسمة بيانات العمود (٥). على ٢ (أى متوسط المتوسطات المتحركة للسنتين المتتاليتين) ويمكن أن نصل إلى نفس النتيجة [بيانات العمود (٦) إذا قمنا بجمع المجموع المتحرك لكل فترتين (العمود (٧) من العمود (٣) ثم بقسمه هذا المجموع على ٨ (عدد سنوات الفترتين) نحصل المتوسطات المتحركة المركزية (العمود (٨) لسنوات السلسلة نفسها. وكما نلاحظ أنها نفس القيم الاتجاهية فى العمود (٦).

وكما ذكرنا سالفاً أن المشكلة الأساسية للمتوسطات المتحركة تتمثل فى اختيار الفترة المناسبة التى يحسب على أساسها المتوسط. ولما كان الهدف من هذا الأسلوب هو تخليص السلسلة الزمنية من الذبذبات والتغيرات حتى يمكن توضيح الاتجاه العام للسلسلة يجب أن تختار - بعناية - الفترة التى تحقق هذا الهدف.

وبصفة عامة، يؤخذ عن طريقة المتوسطات المتحركة أن البيانات فيها تفقد عند بداية ونهاية السلسلة الزمنية. ففى المثال السابق نبدأ العمل ببيانات مكونة من ١١ رقماً وباستخدام متوسط متحرك من الدرجة ٤ فتنتهى بسبعة أرقام. وتبعاً لذلك إذا حسبت المتوسطات المتحركة على أساس عدد قليل من السنوات فإنها تميل إلى تتبع نفس التغيرات فى بيانات السلسلة الأصلية، أى أنها لا تستطيع إزابة التغيرات أو تمهيد الذبذبات الكبيرة. ومن عيوب المتوسطات المتحركة أيضاً أنها تعطى القيم الاتجاهية فقط دون أن تعطى الصيغة أو المعادلة التى يسير عليها التغير، والتى هى أساس التنبؤ. أو بمعنى آخر أن هذه الطريقة لا تبين معدل النمو الحقيقى للظاهرة قد التحليل. ولكن يحسن استخدام هذه الطريقة كأسلوب لتعيين الاتجاه العام إذا لم يكن الاتجاه على شكل خط مستقيم، وعند ما يكون الغرض هو مجرد دراسة حركة السلسلة الزمنية نفسها دون الاهتمام بالتنبؤ بالقيم الاتجاهية لسنوات ليست موجودة أساساً داخل السلسلة الزمنية قيد الدراسة.

٤- طريقة المربعات الصغرى The Least Squares : تعتبر هذه الطريقة التي سبق دراستها فى الفصل السابق (تحليل الانحدار) أكثر الطرق استخداماً لتوفيق معادلة خط الاتجاه العام للبيانات المشاهدة، إذ أنه بواسطتها يمكن تحديد ثوابت الخط المستقيم (الميل م، والجزء المقطوع ج) من واقع بيانات السلسلة الزمنية موضع التحليل، وفى هذه الحالة لن تختلف القيم الاتجاهية المحسوبة من خط إلى آخر نظراً لعدم اختلاف ثوابت الخط المستقيم. كما أن من خصائص هذه الطريقة أنها تجعل مجموع مربعات الفروق بين القيم المشاهدة والقيم الاتجاهية المحسوبة من معادلة الخط المستقيم أقل ما يمكن (راجع الفصل السابق تحليل الانحدار).

وبدأ تعيين الاتجاه العام للسلسلة الزمنية بهذه الطريقة بتمثيل بيانات السلسلة بيانياً وتحديد اتجاهها العام بصفة تقريبية. فإذا أظهر التمثيل البيانى وجود اتجاه عام يأخذ شكل الخط المستقيم كانت معادلته من الدرجة الأولى وصورتها العامة:

$$ص = م س + ج \quad \dots\dots\dots (١-١٣)$$

أما إذا كشف التمثيل البيانى عن وجود اتجاه عام يأخذ شكل منحنى (قطع مكافئ مثلاً) كانت معادلته من الدرجة الثانية وصورتها العامة هى:

$$ص = أ + ب س + ج س^٢ \quad \dots\dots\dots (٢-١٣)$$

حيث أ، ب، ج هى مقادير ثابتة يمكن تقديرها.

وستهتم هنا بتعيين الاتجاه العام على شكل خط مستقيم فقط لما له من أهمية فى دراسة حركة السلسلة الزمنية فى الماضى والتنبؤ الدقيق فيها بتطور الظاهرة قيد التحليل.

ولحساب معاملات أو ثوابت المعادلة (١-١٣) م، ج من البيانات المشاهدة فى السلسلة الزمنية، نستخدم كل من المعادلتين الأساسيتين:

$$م ج ص = م مح س + ن ج \quad \dots\dots\dots (٣-١٣)$$

$$مح س ص = م مح س^٢ + ج مح س \quad \dots\dots\dots (٤-١٣)$$

حيث ن تمثل عدد سنوات للسلسلة الزمنية، ص هى قيمة الظاهرة، س هى

الزمن وتسمى طريقة الحساب هذه «بالطريقة المطولة» وذلك لأن نقطة الأصل بالنسبة للزمن تكون عند السنة الأولى فى السلسلة الزمنية.

مثال:

إذا افترضنا أن الإنتاج السنوى من الفحم (مليون كيلو جرام) فى دولة ما للفترة ١٩٤٨ - ١٩٥٨ تمثل فى السلسلة الزمنية فى الجدول رقم (١٣-٥) ولتقدير أحسن القيم للثوابت م، ج نستخدم بيانات الجدول التالى للتعويض بها فى المعادلتين السابقتين (١٣-٣، ١٣-٤).

جدول رقم (١٣-٨)

حساب خط اتجاه مستقيم بطريقة المربعات الصغرى «الطريقة المطولة»

السنة	ص	س	س ص	س ^٢
١٩٤٨	٥٠,٠	صفر	صفر	صفر
١٩٤٩	٣٦,٥	١	٣٦,٥	١
١٩٥٠	٤٣,٠	٢	٨٦,٠	٤
١٩٥١	٤٤,٥	٣	١٣٣,٥	٩
١٩٥٢	٣٨,٩	٤	١٥٥,٦	١٦
١٩٥٣	٣٨,١	٥	١٩٠,٥	٢٥
١٩٥٤	٣٣,٦	٦	١٩٥,٦	٣٦
١٩٥٥	٣٨,٧	٧	٢٧١,٩	٤٩
١٩٥٦	٤١,٧	٨	٣٣٣,٦	٦٤
١٩٥٧	٤١,١	٩	٣٦٩,٩	٨١
١٩٥٨	٣٣,٨	١٠	٣٣٨,٠	١٠٠
المجموع	٤٣٨,٩	٥٥	٢١١٠,١	٣٨٥

بالتعويض في المعادلة (١٣ - ٣) ينتج أن:

$$٤٣٨,٩ = ٥٥ م + ١١ ج \text{ (حيث } ن = ١١) \text{ (١٣-٥)}$$

وبالتعويض في المعادلة (١٣-٤) ينتج أن:

$$٢١١٠,١ = ٣٨٥ م + ٥٥ ج \text{ (١٣-٦)}$$

ويضرب المعادلة (١٣-٥) في ٥ ينتج أن:

$$٢١٩٤,٥ = ٢٧٥ م + ٥٥ ج \text{ (١٣-٧)}$$

ويطرح المعادلة (١٣-٧) من المعادلة (١٣-٦) ينتج أن:

$$-٨٤,٤ = ١١٠ م \therefore م = \frac{-٨٤,٤}{١١٠} = -٠,٧٦٧$$

وبالتعويض عن قيمة م في المعادلة (١٣-٥) نحصل على قيمة ج

$$٤٣٨,٩ = ٥٥ \times -٠,٧٦٧ + ١١ ج$$

$$٤٣٨,٩ = ٤٢,١٨٥ + ١١ ج$$

$$١١ ج = ٣٩٦,٧١٥ \therefore ج = \frac{٣٩٦,٧١٥}{١١} = ٣٦,٠٦٥$$

وبذلك تكون معاملات أو ثوابت المعادلة (١٣-٥) هي:

$$م = -٠,٧٦٧ \text{ ، } ج = ٣٦,٠٦٥$$

وبالتعويض عن م، ج في معادلة الخط المستقيم (معادلة رقم (١٣-١)، على أساس نقطة أصل سنة ١٩٤٨، ينتج أن:

$$ص = ٧٦٧٠ م + ٣٦,٠٦٥$$

ويمكن تسهيل الحسابات بنقل نقطة الأصل بالنسبة للزمن أو وحدة القياس (س) عند منتصف السلسلة تماماً بحيث يكون عدد قيم الظاهرة التي تسبقها مسار لعدد القيم التي تليها وأخذ انحرافات السنين عن نقطة المنتصف، وبذلك يكون محد س = صفر. وتسمى هذه الطريقة «بالطريقة المختصرة» التي تبحث أيضاً في كيفية إيجاد قيمة كل من م، ج للحصول على معادلة الخط المستقيم، كمايلي:

$$م = \frac{\text{مجمد مس}}{\text{مجمد مس}^2} = \text{عندما مجمد مس} = \text{صفر} \dots\dots\dots (١٣-٨)$$

$$ح = \frac{\text{مجمد مس}}{\text{مجمد ن}} = \text{عندما مجمد مس} = \text{صفر} \dots\dots\dots (١٣-٩)$$

وفى المثال السابق نجد أن السلسلة الزمنية فردية العدد، حيث ن = ١١ سنة، وفى هذه الحالة نختار منتصف هذه السلسلة وهى سنة ١٩٥٣ كنقطة أصل. وإذا أخذنا انحرافات السنين الأخرى فى السلسلة عن هذه السنة نحصل على بيانات سهلة مبسطة للزمن (س) كما يتضح من الجدول التالى:

جدول رقم (١٣-٩)

حساب خط اتجاه مستقيم بطريقة المربعات الصغرى
(الطريقة المختصرة)

السنة	مس	س	مس مس	مس ^٢	القيم الاتجاهية ش
١٩٤٨	٥٠,٠	٥ -	٢٥٠,٠ -	٢٥	٤٣,٧
١٩٤٩	٣٦,٥	٤ -	١٤٦,٠ -	١٦	٤٣,٠
١٩٥٠	٤٣,٠	٣ -	١٢٩,٠ -	٩	٤٢,٢
١٩٥١	٤٤,٥	٢ -	٨٩,٠ -	٤	٤١,٤
١٩٥٢	٢٨,٩	١ -	٣٨,٩ -	١	٤٠,٧
١٩٥٣	٣٨,١	صفر	صفر	صفر	٣٩,٩
١٩٥٤	٣٢,٦	١	٣٢,٦	١	٣٩,١
١٩٥٥	٣٨,٧	٢	٧٧,٤	٤	٣٨,٤
١٩٥٦	٤١,٧	٣	١٢٥,١	٩	٣٧,٦
١٩٥٧	٤١,١	٤	١٦٤,٤	١٦	٣٦,٨
١٩٥٨	٣٣,٨	٥	١٦٩,٠	٢٥	٣٦,١
المجموع	٤٣٨,٩	صفر	٨٤,٤ -	١١٠	

وبالتعويض فى المعادلتين السابقتين (١٣-٨، ١٣-٩) لاستخراج قيمة كل من م، ج نجد أن:

$$٠,٧٦٧ = \frac{٨٤,٤-}{١١٠} = م$$

$$٣٩,٩ = \frac{٤٣٨,٩}{١١} = ح$$

وبالتعويض عن م، ج فى المعادلة الخطية نجد أن:

$$\text{معادلة خط الاتجاه العام هي: } ص = ٠,٧٦٧٠ س + ٣٩,٩ \dots (١٣-١٠)$$

وهى معادلة الخط المستقيم لقيم ص بدلاً من قيم من المختصرة حيث نقطة الأصل سنة ١٩٥٣، س مقاسه بوحدات سنوية (١ سنة)، ص هى كمية الإنتاج السنوى.

ونلاحظ أن قيمة م = ٠,٧٦٧٠ التى حسبناها بالطريقة المختصرة هى نفس القيمة التى حصلنا عليها بالطريقة المطلوبة، بينما قيمة جـ (الجزء المقطوع من الاحداثى الصادى) غير متساوية فى الحالتين، والسبب فى ذلك يرجع إلى أن عملية نقل نقطة الأصل لا تؤثر على ميل الخط المستقيم بقدر أنها تنقل نقطة تقاطع خط الاتجاه العام مع المحور الرأسى فى الرسم، إذ تعد نقطة التقاطع فى الحالة الأخيرة هى نقطة الأصل الجديدة، كذلك نلاحظ أن قيمة جـ = ٣٩,٩ هى القيمة الاتجاهية للسنة التى فى منتصف السلسلة وهى سنة ١٩٥٣ حيث س المختصرة لهذه السنة تساوى صفر وهى نقطة الأصل المتغير الزمن س.

ويمكن رسم خط الاتجاه العام بعد ذلك بمعلومية الميل (م) والجزء المقطوع من الاحداثى الصادى (جـ). ولحقيقة أن كل نقطة تقع على هذا الخط إنما تحقق معادلته، فإنه يمكن معرفة أية قيمة من قيم الظاهرة إذا عرفنا قيمة (س) أى الوحدة الزمنية المناظرة. وكل قيمة من قيم الظاهرة التى تقع على خط الاتجاه العام وتناظر وحدة زمنية معينة يطلق عليها اسم «القيمة الاتجاهية» ويرمز لها بالرمز ص

والقيم الاتجاهية التى فى العمود الأخير فى الجدول رقم (١٣-٩) حصلنا عليها بالتعويض فى المعادلة السابقة حيث أن القيمة الاتجاهية (ص) لسنة ١٩٥٢ مثلاً.

$$ص^{\wedge} = ٠,٧٦٧ \times (١-) + ٣٩,٩$$

$$= ٤٠,٦٦٧ = ٤٠,٧ \text{ تقريباً}$$

وتبعاً للنتائج التى حصلنا عليها يمكن استخدام معادلة خط الاتجاه العام فى التنبؤ بالقيم الاتجاهية لسنوات غير موجودة داخل السلسلة الزمنية نفسها، فمثلاً لحساب القيمة الاتجاهية لإنتاج الفحم فى سنة ١٩٥٩، نعوض فى معادلة خط الاتجاه العام بوضع س = ٦

$$\therefore ص^{\wedge} \text{ لسنة } ١٩٥٩ = ٠,٧٦٧ \times (٦) + ٣٩,٩$$

$$= ٤٠,٦٠٢ \times ٣٩,٩ = ٠٣٥,٢٩٨$$

أما إذا كانت السلسلة الزمنية تحتوى على عدد زوجى من السنين فلا يوجد عندئذ سنة فى منتصف السلسلة تجعل مجموع انحرافات بقية السنين عنها تساوى صفراً. ويمكن التغلب على هذه الصعوبة بأخذ منتصف السلسلة بين السنين فى منتصف السلسلة نقل نقطة الأصل إلى منتصف السلسلة. فإذا كانت السلسلة تقع بين الفترة ١٩٤٨ - ١٩٥٧ فإن منتصف سنة ١٩٥٢ يعتبر نقطة الأصل الجديدة والتى تجعل مح س = صفر إذا جعلنا وحدة الزمن فى هذه الحالة تساوى نصف السنة.

فإذا كانت لدينا سلسلة من ١٠ سنوات فإنه يمكننا توفيق أحسن خط مستقيم للاتجاه العام كالآتى:

جدول رقم (١٣-١٠)
حساب خط اتجاه مستقيم للسلسلة الزوجية بالطريقة المختصرة

السنة	ص	س	س ص	س ٢
١٩٤٨	٥٠,٠	٩-	٤٥٠,٥-	٨١
١٩٤٩	٣٦,٥	٧-	٢٥٥,٥-	٤٩
١٩٥٠	٤٣,٠	٥-	٢١٥,٠-	٢٥
١٩٥١	٤٤,٥	٣-	١٣٣,٥-	٩
١٩٥٢	٢٨,٩	١-	٣٨,٩-	١
١٩٥٣	٢٨,١	١	٣٨,١	١
١٩٥٤	٣٢,٦	٣	٩٧,٨	٩
١٩٥٥	٣٨,٧	٥	١٩٣,٥	٢٥
١٩٥٦	٤١,٧	٨	٢٩١,٩	٤٩
١٩٥٧	٤١,١	٩	٣٦٩,٩	٨١
المجموع	٤٠٥,١	صفر	١٠١,٧-	٣٣٠

من الجدول يتبين أن:

$$\text{محد ص} = ٤٠٥,١ ، \text{محد س ص} = ١٠١,٧- ، \text{محد س} ٢ = ٣٣٠$$

$$\begin{aligned} \text{م} &= \frac{\text{محد س ص}}{\text{محد س} ٢} = \frac{١٠١,٧-}{٣٣٠} \\ \therefore \text{م} &= -٠,٣٠٨ \end{aligned}$$

$$\therefore \text{ج} = \frac{\text{محد ص}}{\text{ن}} = \frac{٤٠٥,١}{١٠}$$

$$\therefore \text{ح} = \frac{٤٠٥,١}{١٠}$$

وبالتعويض عن م، ج في معادلة الخط المستقيم ينتج أن:

$$\hat{ص} = -٠,٣٠٨ س + ٤٠,٥١$$

على أساس أن المتوسط الفرضى هو منتصف سنة ١٩٥٢ وأن الوحدة الزمنية هنا هي نصف السنة.

ولتخليص الظاهرة قيد البحث من أثر الاتجاه العام (مع بقاء وقوعها تحت تأثير العوامل الأخرى وهي أثر الدورة، أثر الموسم، أثر التغيرات العشوائية) نقوم بقسمة القيمة الحقيقية للظاهرة على القيمة الاتجاهية لها (أثر الاتجاه العام). وضربها في ١٠٠ حتى نحصل على أثر الاتجاه العام في شكل نسب مئوية، أى أن:

قيمة الظاهرة بعد تخليصها من أثر الاتجاه العام =

$$\frac{\text{القيمة الحقيقية}}{\text{القيمة الاتجاهية}} \times ١٠٠ = \frac{\hat{ص}}{ص} \times ١٠٠$$

وبالرجوع إلى السلسلة الزمنية الموجودة بالجدول رقم (١٣-٩) فإنه يمكننا تخليص البيانات حصلنا عليها باستخدام المعادلة $\hat{ص} = ٧٦٧٠ س + ٣٩,٩$

بالطريقة المختصرة لطريقة المربعات الصغرى كما هي الحال في الجدول التالى:

ثانياً: تقدير التغيرات الدورية:

تظهر التغيرات الدورية فى أوقات الدورات الاقتصادية، أى يقع تأثيرها فى فترات النمو أو الانكماش الاقتصادى، وتسبب بذلك ذبذبات فى النشاط الاقتصادى. وكما أسفنا أن الغرض الأساسى من تحليل السلسلة الزمنية هو دراسة طبيعية وأسباب الذبذبات بها وعزل التغيرات - ومنها التغيرات الدورية - التى تؤثر فيها عن طريق قياس أثر هذه التغيرات والتنبؤ بوقوعها، لتفادى التأثيرات الخطرة لها ووضع الحلول للصعوبات التى قد تنجم عنها. وتحدد دورة التغيرات الدورية بالفترة بين قمتين أو بين قاعى موجتين متتاليتين على الرسم البيانى للسلسلة الزمنية. وبالرجوع إلى بيانات الجدول رقم (١٣-١١) والشكل رقم (١٣-٨) يمكن

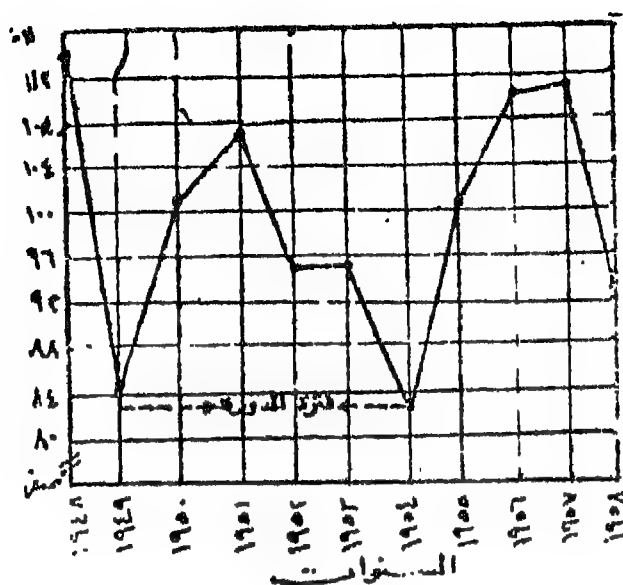
تحديد فترة الدورة بالمسافة الأفقية بين القاعين الممثلين للسنتين ١٩٤٩، ١٩٥٤، وهي تمثل دورة طولها خمس سنوات.

جدول رقم (١٣-١١)

تخليص بيانات إنتاج الفحم (مليون كيلو جرام) من أثر الاتجاه العام

السنة	ص	س	س
١٩٤٨	٥٠,٠	٤٣,٧	١١٤,٤٢
١٩٤٩	٣٦,٥	٤٣,٠	٨٤,٨٨
١٩٥٠	٤٣,٠	٤٢,٢	١٠١,٨٩
١٩٥١	٤٤,٥	٤١,٤	١٠٧,٤٩
١٩٥٢	٣٨,٩	٤٠,٧	٩٥,٥٨
١٩٥٣	٣٨,١	٣٩,٩	٩٥,٤٩
١٩٥٤	٣٢,٦	٣٩,١	٨٣,٣٧
١٩٥٥	٣٨,٧	٣٨,٤	١٠٠,٧٨
١٩٥٦	٤١,٧	٣٧,٦	١١٠,٩٠
١٩٥٧	٤١,١	٣٦,٨	١١١,٦٨
١٩٥٨	٣٣,٨	٣٦,١	٩٣,٦٣

ويكتشف عملية تقرير التغيرات الدورية صعوبة رئيسية وهي أن التأثيرات الدورية لا تعيد نفسها إذ أنها تقع كل فترة طويلة من الزمن وبطريقة غير منتظمة عكس التغيرات الموسمية. ولكن يمكن القول أن دراسة الدورات في سلسلة زمنية معينة تعبر عن نموذج عام تتشابه فيه الدورات جميعها مع وجود بعض الفروق بينها في طول الفترة والسعة. وبناء على ذلك فإنه عند تحليل السلسلة الزمنية لتقدير التأثيرات الدورية يجب أن يتحدد ما إذا كانت السلسلة الزمنية سنوية أم موسمية.



شكل رقم (١٣-٨) تحديد فترة الدورية

ففى حالة إذا كانت السلسلة الزمنية سنوية فإن قيمة الظاهرة - بعد إهمال تأثير التغيرات العشوائية - تكون تحت تأثير عاملين فقط هما: أثر الاتجاه العام وأثر التغيرات الدورية، لأن العامل الثالث وهو أثر التغيرات الموسمية لا يوجد له تأثير إذا كانت السلسلة سنوية وعلى ذلك فإنه يمكن تقدير أثر التغيرات الدورية التى يمكن اعتبارها انحرافات عن القيم الاتجاهية السنوية عن طريق قسمة القيم الأصلية الظاهرة على القيم الاتجاهية لها، إذ أن:

قيمة الظاهرة (ص) = أثر الاتجاه العام (ص) × أثر التغيرات الدورية (د) [مع إهمال أثر التغيرات العشوائية].

فتكون التغيرات الدورية (د) = $\frac{\text{قيمة الظاهرة (ص)}}{\text{القيمة الاتجاهية (ص)}}$ ويفضل ضرب هذه

النسبة في ١٠٠ حتى نحصل على النسبة المئوية للتغيرات الدورية في السلسلة الزمنية موضع التحليل.

وباستخدام بيانات الجدول رقم (١٣-١٢) تمكنا من حساب أحسن خط مستقيم يمثل الاتجاه العام للسلسلة الزمنية التي تبين إنتاج محصول القمح (ألف طن) في دولة ما في الفترة من ١٩٤٦ - ١٩٥٨ بطريقة المربعات الصغرى، والتي أعطتنا المعادلة الآتية (على أساس نقطة الأصل هي سنة ١٩٥٢ والوحدة الزمنية هي سنة):

$$\hat{ص} = ١,٨٢ س + ٩٢,٤$$

وباستخدام هذه المعادلة حصلنا على القيم الاتجاهية ($\hat{ص}$) لسنوات السلسلة عن طريق التعويض بقيم ($س$) المناظرة وهي الفترة ١٩٤٦ - ١٩٥٨. وبقيمة القيم الأصلية للظاهرة على القيم الاتجاهية حصلنا على قيمة الظاهرة بعد تخليص البيانات الأصلية من أثر الاتجاه العام، وهذا يعنى بالتالى أثر التغيرات الدورية التي وضعت في صورة نسب مئوية، حتى تسهل معها عملية المقارنة، ووضعت النتائج في الجدول التالي:

أما إذا كانت السلسلة الزمنية موسمية فإنه بإهمال أثر التغيرات فإن قيمة لظاهرة ($ص$) تكون حيث:

$$ص = \hat{ص} \times و \times د$$

وتكون التغيرات الدورية ($د$) في هذه الحالة هي:

$$د = \frac{ص}{\hat{ص} \times و}$$

ويعنى ذلك أننا أردنا تقدير أثر التغيرات الدورية فإننا نقوم بقسمة القيم الأصلية للظاهرة على حاصل ضرب القيم الاتجاهية لها في النسب الموسمية مع إهمال التغيرات العشوائية وبذلك نحصل على مقياس للذبذبات الدورية.

جدول رقم (١٣-١٢)
حساب أثر التغيرات الدورية لإنتاج القمح (ألف طن) في دولة ما
في الفترة من ١٩٤٦ - ١٩٥٨

السنة	الإنتاج (ص)	القيم الاتجاهية (ص)	النسبة المئوية للتغيرات الدورية	السنة	الإنتاج (ص)	القيم الاتجاهية (ص)	النسبة المئوية للتغيرات الدورية
١٩٤٦	٨٢,٣	٨١,٥	١٠١,٠	١٩٥٣	٩٦,١	٩٤,٢	١٠٢,٠
١٩٤٧	٩٥,٦	٨٣,٣	١١٤,٨	١٩٥٤	٨٩,١	٩٦,٠	٩٢,٨
١٩٤٨	٨١,٣	٨٥,١	٩٥,٥	١٩٥٥	١١٢,٢	٩٧,٩	١١٤,٦
١٩٤٩	٦٩,٤	٨٧,٠	٧٨,٨	١٩٥٦	١٠٠,٤	٩٩,٧	١٠٠,٧
١٩٥٠	٩٢,٨	٨٨,٩	١٠٤,٤	١٩٥٧	١٠٦,٩	١٠١,٥	١٠٥,٣
١٩٥١	٨٣,٤	٩٠,٦	٩٢,١	١٩٥٨	١٩٦,٦	١٠٣,٣	٩٣,٥
١٩٥٢	٨٨,٢	٩٢,٤	٩٥,٥				

وجدير بالذكر أن العديد من الباحثين الاحصائيين ينصح في مثل هذه الحالات بأخذ سلسلة زمنية طويلة للظاهرة قيد التحليل، حتى يمكن أن تظهر ملامح التأثيرات الدورية بوضوح، وفي الوقت نفسه لا يجب إهمال التغيرات العشوائية على الرغم من صعوبة تحديدها أو التنبؤ بها.

ثالثاً: التغيرات الموسمية:

ذكرنا أن التغيرات الموسمية هي إحدى العوامل التي يظهر أثرها على المنحنى التاريخي للسلسلة الزمنية، وأن سبب هذه التغيرات يرجع إلى عوامل طبيعية مثل الظروف المناخية أو الحركات التكتونية السريعة (الزلازل والبراكين)، وعوامل إجتماعية كالعادات والتقاليد والمواسم والأعياد التي تتكرر كل عام. وتفيد دراسة التغيرات الموسمية في معرفة نموذج التغيرات نفسها وعمل المقارنات بين التغيرات

الموسمية في السنوات المختلفة، ومعرفة إلى أى مدى تؤثر هذه التغيرات في قيم الظاهرة.

ولحساب أثر التغيرات الموسمية هناك طرق عديدة الغرض الأساسى منها هو معرفة قيمة هذا الأثر على السلسلة الزمنية بعد استبعاد كل العوامل الأخرى المؤثرة، أو بمعنى آخر تحديد المعامل الموسمى كما لو أن قيم الظاهرة لم تتأثر إلا بالتأثير الموسمى فقط. ويطلق على مجموعة الأرقام التى توضح القيم النسبية الموسمية للظاهرة خلال شهور أو فصول السنة إسم «الدليل الموسمى Seasonal Index»، للظاهرة. وسنعرض هنا باختصار لطريقتين من طرق حساب الدليل الموسمى وهما: طريقة متوسط النسب المئوية الموسمية، وطريقة النسبة المئوية للاتجاه العام (أو نسبة الاتجاه العام).

١ - طريقة متوسط النسب المئوية: تستخدم هذه الطريقة فى تقدير التغيرات الموسمية حيث يعبر فيها عن بيانات كل موسم (سواء كان الموسم يوماً أو أسبوعاً، أو شهر أو فصلاً) خلال السنة كنسبة مئوية من المتوسط فى السنة. ثم نحصل على متوسط النسبة للمواسم المتقابلة فى مختلف السنوات وذلك باستخدام مقياس المتوسط الحسابى أو الوسيط (من الأفضل حذف القيم المتطرفة أو الشاذة عند استخدام المتوسط الحسابى) وتمثل - فى النهاية - النسب المئوية للمواسم «الدليل الموسمى» المطلوب. فإذا كان متوسط الدليل الموسمى أكبر من أو أقل من ١٠٠٪ فيجب تعديله بالضرب فى معامل ملائم.

مثال:

يأخذ سلسلة زمنية توضح الطاقة الكهربائية الشهرية (مليون كيلو واط ساعة) المستهلكة فى إضاءة الشوارع والطرق فى دولة ما فى السنوات ١٩٦١ - ١٩٥٨ وجد أن توزيع هذه الطاقة حسب شهور السنة كالآتى:

جدول رقم (١٣-١٣)
الطاقة الكهربائية (مليون كيلو وات ساعة).
المستهلكة في دولة ما في الفترة من ١٩٥١ - ١٩٥٨

الشهر السنة	١٩٥١	١٩٥٢	١٩٥٣	١٩٥٤	١٩٥٥	١٩٥٦	١٩٥٧	١٩٥٨
يناير	٣١٨	٣٤٢	٣٦٧	٣٩٢	٤٢٠	٤٥٣	٤٨٧	٥٢٩
فبراير	٢٨١	٣٠٩	٣٢٨	٣٤٩	٣٧٨	٤١٢	٤٤٠	٤٧٧
مارس	٢٧٨	٢٩٩	٣٢٠	٣٤٢	٣٧٠	٣٩٨	٤٢٩	٤٦٣
أبريل	٢٥٠	٢٦٨	٢٨٧	٣١١	٢٣٤	٢٦٢	٢٩٣	٤٢٣
مايو	٢٣١	٢٤٩	٢٦٩	٢٩٠	٣١٤	٣٤١	٣٧٠	٣٩٨
يونيو	٢١٦	٢٣٦	٢٥١	٢٧٣	٢٩٦	٣٢٢	٣٤٧	٣٨٠
يوليو	٢٢٣	٢٤٢	٢٥٩	٢٨٢	٣٠٥	٣٣٥	٣٥٧	٣٨٩
أغسطس	٢٤٥	٢٦٢	٢٨٤	٣٠٥	٣٣٠	٣٥٩	٣٨٨	٤١٩
سبتمبر	٢٦٩	٢٨٨	٣٠٩	٣٢٨	٣٥٦	٣٩٢	٤١٥	٤٤٨
أكتوبر	٣٠٢	٣٢١	٣٤٥	٣٦٤	٣٩٦	٤٢٧	٤٥٧	٤٩٣
نوفمبر	٣٢٥	٣٤٢	٣٦٧	٣٨٩	٤٢٢	٤٥٤	٤٩١	٥٢٦
ديسمبر	٣٤٧	٣٦٤	٣٩٤	٤١٧	٤٥٢	٤٨٣	٥١٦	٥٦٠
المجموع	٢٢٨٥	٣٥٢٢	٣٧٨٠	٤٠٤٢	٤٣٧٣	٤٧٣٨	٥٠٩٠	٥٥٠٥
المتوسط	٢٧٣,٧	٢٩٣,٥	٣١٥٠	٣٣٦٠,٨	٣٦٤٠,٤	٢٩٤,٨	٤٢٤,٢	٤٥٨,٧

ولتقدير الدليل الموسمي بطريقة متوسط النسب المئوية بقسمة البيانات الشهرية المعطاه على المتوسطات الشهرية المقابلة لكل سنة مع التغير عن النتيجة كنسبة مئوية فنحصل على البيانات التي يوضحها الجدول التالي:

جدول رقم (١٣-١٦)
النسب المئوية والدليل الموسمي للطاقة الكهربائية المستهلكة
في دولة ما في الفترة من ١٩٥١ - ١٩٥٨

الشهر السنة	١٩٥١	١٩٥٢	١٩٥٣	١٩٥٤	١٩٥٥	١٩٥٦	١٩٥٧	١٩٥٨	المجموع	المتوسط
يناير	١١٦,٢	١١٦,٥	١١٦,٥	١١٦,٤	١١٥,٣	١١٤,٧	١١٤,٨	١١٥,٣	٩٢٥,٧	١١٥,٧
فبراير	١٠٢,٧	١٠٥,٣	١٠٤,١	١٠٣,٦	١٠٣,٧	١٠٤,٤	١٨٣,٧	١٠٤,٠	٨٣١,٥	١٠٣,٩
مارس	١٠١,٦	١٠١,٩	١٠١,٦	١٠١,٥	١٠١,٥	١٠٠,٨	١٠١,١	١٠٠,٩	٨١٠,٩	١٠١,٤
أبريل	٩١,٣	٩١,٣	٩١,١	٩٢,٣	٩١,٧	٩١,٧	٩٢,٦	٩٢,٢	٨٢٤,٢	٩١,٨
مايو	٨٤,٤	٨٤,٨	٨٥,٤	٨٦,١	٨٦,٢	٨٦,٤	٨٧,٢	٨٦,٨	٦٨٧,٣	٨٥,٩
يونيو	٧٨,٩	٨٠,٤	٧٩,٧	٨١,١	٨٢,٢	٨١,٦	٦١,٨	٨٢,٨	٦٤٧,٥	٨٠,٩
يوليو	٨١,٥	٨٢,٥	٨٢,٢	٨٣,٧	٨٣,٧	٨٤,٩	٨٤,٢	٨٤,٨	٦٦٧,٥	٨٣,٤
أغسطس	٨٩,٥	٨٩,٣	٩٠,٢	٩٠,٦	٩٠,٦	٩٠,٩	٩١,٥	٩١,٢	٧٢٣,٩	٩٠,٥
سبتمبر	٩٨,٣	٩٨,١	٩٨,١	٩٧,٤	٩٧,٧	٩٩,٣	٩٧,٨	٩٧,٧	٧٨٤,٤	٩٨,١
أكتوبر	١١٠,٣	١٠٩,٤	١٠٩,٥	١٠٨,١	١٠٨,٧	١٠٨,٢	١٠٧,٧	١٠٧,٥	٨٦٩,٤	١٠٨,٧
نوفمبر	١١٨,٧	١١٦,٥	١١٦,٥	١١٥,٥	١١٥,٨	١١٥,٠	١١٥,٧	١١٤,٧	٩٢٨,٤	١١٦,١
ديسمبر	١٢٦,٨	١٢٤,٠	١٢٥,١	١٢٣,٨	١٢٤,٠	١٢٢,٣	١٢١,٦	١٢٢,١	٩٨٩,٧	١٢٣,٧

فعلى سبيل المثال القيمة الأولى يناير ١٩٥١ في الجدول السابق تحسب كمايلي:

$$١١٦,٢ = ١٠٠ \times \frac{٣١٨}{٢٧٣,٦} \dots \text{وهكذا لباقي شهور نفس السنة.}$$

وجدير بالذكر أنه إذا كانت الشهور كلها متساوية من حيث طاقة الاضاءة المستخدمة لانعدمت التغيرات الرسمية، ولو وجدنا أن متوسط كل شهر يساوى المتوسط العام ولكانت نسبة موسمية تساوى ١٠٠٪. وحيث أننا وجدنا أن النسبة الموسمية للشهر الأول هي ١١٥,٧ فمعنى هذا أن متوسط الطاقة الكهربائية

المستخدمة فى هذا الشهر يزيد عن المتوسط العام بمقدار ١٥,٧ ٪، بينما نجد أن النسبة الموسمية لشهر يوليو هى ٨٠,٩ ٪ ومعنى هذا أن متوسط الطاقة الكهربائية المستخدمة فى هذا الشهر ينقص عن المتوسط العام بمقدار ١٩,١ ٪، ولو جمعنا الزيادة والنقص فى النسب المئوية وكان الناتج أكبر من الصفر فيجب تعديل النسب المئوية للشهور بضربها فى معامل ملائم. وبما أن مجموع متوسط النسب المئوية لكل الشهور هو ١٢٠,١ ٪ وهو قريب جداً من المجموع المطلوب ١٢٠ ٪ بمتوسط ١٠٠ ٪ لشهور السنة جميعها، فإنه ليس من الضرورى تعديل النسب المئوية للشهور والتي تعبر فى هذه الحالة عن «الدليل الموسمى» المطلوب.

ويمكن إستخدام طريقة أخرى لحساب متوسط النسب المئوية تقوم على أساس حساب متوسط عام (متوسط المتوسطات) للمواسم - سواء كانت على أساس يومى أو أسبوعى أو شهرى ... الخ - عن طرق حساب متوسط قيم كل موسم من مواسم السلسلة الزمنية موضع التحليل، ثم بعد ذلك يحسب المتوسط الحسابى لمتوسطات المواسم وهو ما يعبر عنه بالمتوسط العام. ويقسم المتوسط الحسابى لكل موسم على المتوسط العام وضرب الناتج فى ١٠٠ نحصل على نتائج فى شكل نسب مئوية هى ما يطلق عليه اسم «النسب الموسمية». ويمكن توضيح ذلك بالصورة الرمزية الآتية:

السنة (١)	السنة (٢)	السنة (٣)	السنة (٤)	
١١س	٢١س	٣١س	٤١س	الموسم (١)
١٢س	٢٢س	٣٢س	٤٢س	الموسم (٢)
١٣س	٢٣س	٣٣س	٤٣س	الموسم (٣)

ويكون:

$$\begin{aligned} \bar{1}س &= \frac{١١س + ٢١س + ٣١س + ٤١س}{٤} = \text{متوسط الموسم الأول للسنوات الأربع} \\ \bar{2}س &= \frac{١٢س + ٢٢س + ٣٢س + ٤٢س}{٤} = \text{متوسط الموسم الأول للسنوات الأربع} \end{aligned}$$

$$\bar{r}_3 = \frac{r_{31} + r_{32} + r_{33} + r_{34}}{4} = \text{متوسط الموسم الثالث للسنوات الأربع}$$

$$\bar{r} = \frac{\bar{r}_1 + \bar{r}_2 + \bar{r}_3}{3} = \text{والمتوسط العام للمواسم}$$

وعليه فإن:

$$\text{نسبة الموسم الأول} = \frac{\bar{r}_1}{\bar{r}} \times 100, \text{ نسبة الموسم الثاني} = \frac{\bar{r}_2}{\bar{r}} \times 100$$

$$\text{نسبة الموسم الثالث} = \frac{\bar{r}_3}{\bar{r}} \times 100$$

وتمثل النسب المثوية للمواسم «الدليل الموسمي» للسلسلة الزمنية

وإذا طبقت هذه الطريقة على بيانات الجداول رقم (١٣-١٣) فإننا نحصل على نسب مثوية مساوية تقريباً للنسب المثوية في الجدول رقم (١٣-١٤)، وهو ما يوضحه الجدول التالي:

جدول رقم (١٣-١٥)

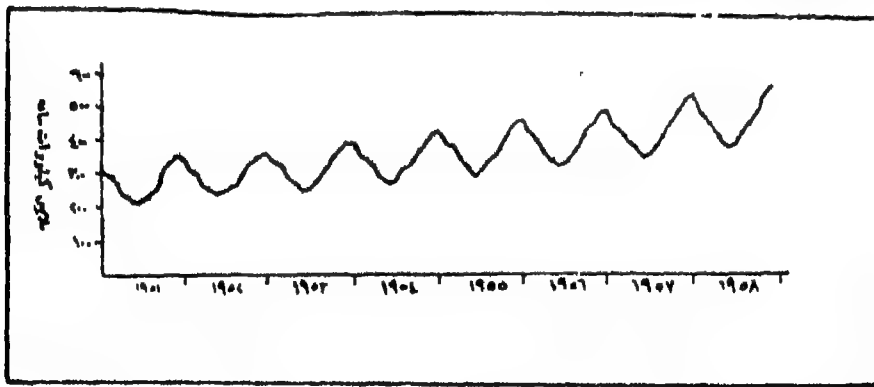
حساب النسب المثوية الموسمية بطريقة المتوسط العام لبيانات

الطاقة الكهربائية بالجدول (١٣-١٣)

الشهر	المجموع	المتوسط	النسبة الموسمية	الشهر	المجموع	المتوسط	النسبة الموسمية
يناير	٣٣٠٨	٤١٣,٥	١١٥,٦	أغسطس	٢٥٩٢	٣٢٤,٠	٩٠,٦
فبراير	٢٩٧٤	٣٧١,٧	١٠٣,٩	سبتمبر	٢٨٠٥	٣٥٠,٦	٩٨,٨
مارس	٢٨٩٩	٣٦٢,٤	١٠١,٣	أكتوبر	٣١٠٥	٣٨٨,١	١٠٨,٥
أبريل	٢٦٢٨	٣٢٨,٥	٩١,٩	نوفمبر	٣٣١٦	٤١٤,٥	١١٥,٩
مايو	٢٤٦٢	٣٠٧,٧	٨٦,٠	ديسمبر	٣٥٣٣	٤٤١,٦	١٢٣,٥
يونيو	٢٣٢١	٢٩٠,١	٨١,١	المجموع		٤٢٩١,٧	
يوليو	٢٣٩٢	٢٩٩,٠	٨٣,٦	المتوسط العام		٣٥٧,٦	

٢- طريقة النسبة المئوية للاتجاه العام: تهدف هذه الطريقة إلى استبعاد التغيرات الدورية والعشوائية من البيانات. وكما لاحظنا أن معظم بيانات الظواهر تتأثر بالاتجاه العام لذلك يجب استبعاد هذا الأثر أيضاً قبل تقدير التأثير الموسمي. ويعبر في هذه الطريقة عن بيانات الموسم كنسبة مئوية من القيم الاتجاهية في الموسم. وباستخدام متوسط ملائم لهذه النسب للمواسم المتقابلة نحصل على الدليل الموسمي المطلوب. وكما في طريقة متوسط النسب المئوية تعدل النسب المئوية التي نحصل عليها إذا لم يكن متوسطها ١٠٠٪.

فمثلاً إذا استخدمنا المتوسطات الشهرية المقابلة لكل سنة من سنوات السلسلة الزمنية (٥١ - ١٩٥٨) في الجدول رقم (١٣-١٣) للحصول على الدليل الموسمي باستخدام طريقة النسبة المئوية للاتجاه العام، يجب استخدام طريقة المربعات الصغرى - السابق شرحها - للحصول على القيم الاتجاهية. فمن الرسم البياني (شكل رقم ١٣-٩) للبيانات الأصلية يتضح أن الاتجاه العام طويل المدى يمكن تقريبه بصورة مناسبة بخط مستقيم.



شكل رقم (١٣-٩)

تطور استخدام الطاقة الكهربائية (مليون كيلووات ساعة)
في دولة ما في الفترة ١٩٥٨ - ١٩٥١

وبافتراض أن القيم الشهرية في الجدول رقم (١٣-١٣) تقابل منتصف الشهر فإن المتوسطات السنوية في هذا الجدول تقابل ٣٠ يونيو أو ١ يوليو للسنة القابلة

لكل متوسط، وبأخذ وحدة الزمن في هذه الحالة نصف سنة، ونقطة الأصل هي ٣١ ديسمبر ١٩٥٤، أو يناير ١٩٥٥، فإنه يمكن توفيق أحسن خط مستقيم للاتجاه العام كالآتي:

جدول رقم (١٣-١٦)

حساب خط الاتجاه العام بطريقة المربعات الصغرى (الطريقة المختصرة)

السنة	ص	س	س ص	س ^٢
١٩٥١	٢٧٣,٧	٧-	١٩١٥,٩-	٤٩
١٩٥٢	٢٩٣,٥	٥-	١٤٦٧,٥-	٢٥
١٩٥٣	٣١٥,٠	٣-	٩٤٥,٠-	٩
١٩٥٤	٣٣٦,٨	١-	٣٣٦,٨-	١
١٩٥٥	٣٦٤,٤	١	٣٦٤,٤	١
١٩٥٦	٣٩٤,٨	١	١١٨٤,٤	٩
١٩٥٧	٤٢٤,٢	٥	٢١٢١,٠	٢٥
١٩٥٨	٤٥٨,٧	٧	٣٢١٠,٩	٤٩
المجموع	٢٨٦١,١	صفر	٢٢١٥,٥	١٦٨

وحيث أن معادلة الخط المستقيم هي $ص = م س + ح$ ،

وبالتعويض عن م بالقيمة $\left(\frac{\text{مجم ص}}{\text{مجم س}}\right)$ ، ح بالقيمة $\left(\frac{\text{مجم ص}}{\text{ن}}\right)$ ينتج أن:

$$ص = \left(\frac{\text{مجم ص}}{\text{مجم س}}\right) س + \left(\frac{\text{مجم ص}}{\text{ن}}\right)$$

$$= ١٣,١٨٨ س + ٣٥٧,٦$$

ومن هذه المعادلة نستنتج أن قيم (ص) تزيد ١٣,١٨٨ كل نصف سنة أو ٢,٢ كل شهر (١٣,١٨٨ ÷ ٦)، وبالتالي فعند س = صفر (١ يناير ١٩٥٥) فإن

ص = ٣٥٧,٦ ، وبعد نصف شهر (١٥ يناير ١٩٥٥) فإن قيمة (ص) تصبح ٣٥٨,٧ $\times ٢,٢ \times \frac{1}{4}$ (٣٥٧,٦ +) وهي القيمة الاتجاهية (ص) المقابلة لشهر يناير ١٩٥٥. وبإضافة المتتالية ٢,٢ إلى ٣٥٨,٧ نحصل على القيمة الاتجاهية لشهر فبراير ١٩٥٥ وهي ٣٦٠,٩ (٣٥٨,٧ + ٢,٢). ولشهر مارس ١٩٥٥ تكون القيمة الاتجاهية (ص) هي ٣٦٣,١ (٣٦٠,٩ + ٢,٢)، وهكذا. وبنفس الطريقة فإنه بالطرح المتتالي للقيمة ٢,٢ من القيمة ٣٥٨,٧ فإننا نحصل على القيمة الاتجاهية لشهر ديسمبر ١٩٥٤ ونوفمبر ١٩٥٤ وهي: ٣٥٦,٥ (٣٥٨,٧ - ٢,٢) و ٣٥٤,٣ (٣٥٦,٥ - ٢,٢) على الترتيب. وبصورة مشابهة نحصل على القيم الاتجاهية الشهرية الموضحة بالجدول التالي:

جدول رقم (١٣-١٧)

القيم الاتجاهية لتطور الطاقة الكهربائية

المستهلكة في دولة ما في الفترة من ١٩٥١ - ١٩٥٨

الشهر السنة	١٩٥١	١٩٥٢	١٩٥٣	١٩٥٤	١٩٥٥	١٩٥٦	١٩٥٧	١٩٥٨
يناير	٢٥٣,١	٢٧٩,٥	٣٠٥,٩	٣٣٢,٣	٣٥٨,٧	٣٨٥,١	٤١١,٥	٤٣٧,٩
فبراير	٢٥٥,٣	٢٨١,٧	٣٠٨,١	٣٣٤,٧	٣٦٠,٩	٣٨٧,٣	٤١٣,٧	٤٤٠,١
مارس	٢٥٧,٥	٢٨٣,٩	٣١٠,٣	٣٣٦,٧	٣٦٣,١	٣٨٩,٥	٤١٥,٩	٤٤٢,٣
أبريل	٢٥٩,٧	٢٨٦,١	٣١٢,٥	٣٣٨,٩	٣٦٥,٣	٣٩١,٧	٤١٨,١	٤٤٤,٥
مايو	٢٦١,٩	٢٨٨,٣	٣١٤,٧	٣٤١,١	٣٦٧,٥	٣٩٣,٩	٤٢٠,٣	٤٤٦,٧
يونيو	٢٦٤,١	٣٩,٥	٣١٦,٩	٣٤٣,٣	٣٦٩,٧	٣٩٦,١	٤٢٢,٥	٤٤٨,٩
يوليو	٢٦٦,٣	٢٩٢,٧	٣١٩,١	٣٤٥,٥	٣٧١,٩	٣٩٨,٣	٤٢٤,٧	٤٥١,١
أغسطس	٢٦٨,٥	٢٩٤,٩	٣٢١,٣	٣٤٧,٧	٣٧٤,١	٤٠٠,٥	٤٢٦,٩	٤٥٣,٣
سبتمبر	٢٧٠,٧	٢٩٧,١	٣٢٣,٥	٣٤٩,٩	٣٧٦,٣	٤٠٢,٧	٤٢٩,١	٤٥٥,٥
أكتوبر	٢٧٢,٩	٢٩٩,٣	٣٢٥,٧	٣٥٢,١	٣٧٨,٥	٤٠٤,٩	٤٣١,٣	٤٥٧,٥
نوفمبر	٢٧٥,١	٣٠١,٥	٣٢٧,٩	٣٥٤,٣	٣٨٠,٧	٤٠٧,١	٤٣٣,٥	٤٥٩,٩
ديسمبر	٢٧٧,٣	٣٠٣,٧	٣٣٠,١	٣٥٦,٥	٣٨٢,٩	٤٠٩,٣	٤٣٥,٧	٤٦٢,١

وبقسمة كل قيمة من القيم الشهرية الأصلية بالجدول رقم (١٣-١٣) على القيمة الاتجاهية بالجدول السابق (جدول رقم ١٣-١٧) وبضربها في ١٠٠ نحصل على نتائج في شكل نسب مئوية. فعلى سبيل المثال تحسب القيمة الأولى بالجدول كالآتي $100 \times \frac{318}{253,1} = 125,6\%$ وبالمثل يمكن حساب بقية القيم فنحصل على القيم الموضحة بالجدول التالي:

جدول رقم (١٣-١٨)

النسب الموسمية للطاقة الكهربائية المستهلكة في دولة ما
في الفترة من ١٩٥١ - ١٩٥٨

الشهر السنة	١٩٥١	١٩٥٢	١٩٥٣	١٩٥٤	١٩٥٥	١٩٥٦	١٩٥٧	١٩٥٨
يناير	١٢٥,٦	١٢٢,٤	١٢٠,٠	١١٨,٠	١١٧,١	١١٧,٦	١١٨,٣	١٢٠,٨
فبراير	١١٠,١	١١٠,٠	١٠٦,٥	١٠٤,٣	١٠٤,٧	١٠٦,٤	١٠٦,٤	١٠٨,٤
مارس	١٠٨,٠	١٠٥,٣	١٠٣,١	١٠١,٦	١٠١,٩	١٠٢,٢	١٠٣,١	١٠٤,٧
أبريل	٩٦,٣	٩٣,٧	٩١,٨	٩١,٨	٩١,٤	٩٢,٤	٩٤,٠	٩٥,٢
مايو	٨٨,٢	٨٦,٤	٨٥,٥	٨٥,٠	٨٥,٢	٨٦,٦	٨٨,٠	٨٩,١
يونيو	٨١,٨	٨١,٢	٨٩,٢	٧٩,٥	٨٠,١	٨١,٣	٨٢,١	٨٤,٧
يوليو	٨٣,٧	٨٢,٧	٨١,٢	٨١,٦	٨٢,٠	٨٤,١	٨٤,١	٨٦,٢
أغسطس	٩١,٢	٨٨,٨	٨٨,٤	٨٧,٧	٨٨,٢	٨٩,٦	٩٠,٩	٩٢,٤
سبتمبر	٩٩,٤	٩٦,٩	٩٥,٥	٩٣,٧	٩٤,٦	٩٧,٣	٩٦,٧	٩٨,٤
أكتوبر	١١٠,٧	١٠٧,٣	١٠٥,٩	١٠٣,٤	١٠٤,٦	١٠٥,٥	١٠٦,٠	١٠٧,٧
نوفمبر	١١٨,١	١١٣,٤	١١١,٩	١٠٩,٨	١١٠,٨	١١١,٥	١١٣,٣	١١٤,٤
ديسمبر	١٢٥,١	١١٩,٩	١١٩,٤	١١٧,٠	١١٨,٠	١١٨,٠	١١٨,٤	١٢١,٢

وإذا قورنت المتوسطات الشهرية ببعض في السلسلة للزمنية قيد التحليل فإن الذبذبات في هذه المتوسطات الشهرية ترجع إلى التغيرات الموسمية لأن التغيرات

الأخرى تستبعد تلقائياً خصوصاً إذا كان عدد سنين السلسلة كبيراً. ويجب استبعاد القيم المتطرفة فى السلسلة الزمنية والتي تكون بسبب الظواهر الفجائية، ولذلك يفضل استخدام الوسيط بدلاً من المتوسط الحسابى، لأن الوسيط لا يتأثر بالقيم المتطرفة. وفى هذه الحالة ترتب النسب للسنتين فى المنتصف (راجع كيفية حساب الوسيط فى حالة التوزيعات غير المبوية ذات الأعداد الزوجية من القيم). وأخيراً تنسب وسيط كل شهر إلى الوسيط الشهرى العام (وهو عبارة عن مجموع الوسيطات الشهرية ١١٩٦,١ مقسوماً على ١٢، أى يساوى ١٩٩,٧٪). ونظراً لأن مجموع قيم الوسيطات الشهرية وهو ١١٩٦,١ أقل من المجموع الذى يجب أن يكون وهو ١٢٠٠ فإننا نعدل هذه القيم بضربها فى النسبة $1200 \div 1196,1$. وبهذه الطريقة نحصل على الدليل الموسمى المطلوب كما هو موضح بالجدول التالى:

جدول رقم (١٣-١٩)

الوسيط والدليل الموسمى للبيانات السابق عرضها فى الجدول رقم (١٣-١٨)

الشهر	الوسيط	الدليل الموسمى	الشهر	الوسيط	الدليل الموسمى	الشهر	الوسيط	الدليل الموسمى
يناير	١١٩,٢	١١٩,٦	مايو	٨٦,٥	٨٦,٨	سبتمبر	٩٦,٨	٩٧,١
فبراير	١٠٦,٤	١٠٦,٧	يونيو	٨١,٢	٨١,٥	أكتوبر	١٠٦,٠	١٠٦,٣
مارس	١٠٣,٤	١٠٣,٤	يوليو	٨٣,٢	٨٣,٥	نوفمبر	١١٢,٦	١١٣,٠
أبريل	٩٣,٠	٩٣,٣	أغسطس	٨٩,٢	٨٩,٥	ديسمبر	١١٨,٩	١١٩,٣

ومما تجدر ملاحظته على الجدول السابق أن أرقام الدليل الموسمى فى الأشهر السبعة الأولى أكبر من نظائرها التى حصلنا عليها بطريقة متوسط النسب المثوية (راجع الجدول رقم (١٣-١٤)، بينما تكون أرقام الدليل الموسمى فى الأشهر الخمسة الأخيرة أقل من مثيلتها فى الجدول رقم (١٣-١٤).

ومن أهم مزايا قياس التغيرات الموسمية هو استخدامها في التنبؤ بمقدار التأثيرات الموسمية في السنوات اللاحقة لسنوات السلسلة الزمنية قيد التحليل. فإذا ما أمكن التنبؤ بالقيمة الاتجاهية في سنة معينة من سنوات السلسلة، فإنه يمكن التنبؤ بقيمة كل موسم في تلك السنة على حدة إذا كانت النسب المئوية محسوبة. ولنفرض أننا نريد أن نتنبأ بمتوسط الطاقة الكهربائية المستخدمة في الاضائة لشهور سنة ١٩٥٩ باستخدام المعادلة الخطية $ص = ١٣,٨٨ س + ٣٥٧,٦$ حيث $س$ تمثل وحدة الزمن وهي نصف السنة. وبطريقة إضافة المتتالية ٢,٢ (المعدل الشهري الذي حصلنا عليه من قسمة المعدل نصف السنوي ١٣,١٨٨ على ٦) إلى القيمة الاتجاهية لشهر ديسمبر ١٩٥٨ وهي ٤٦٢,١ نحصل على القيمة الاتجاهية لشهر يناير ١٩٥٩ وهي ٤٦٤,٣ (٤٦,١٠ + ٢,٢)، وبإضافة المتتالية إلى القيمة الاتجاهية الأخيرة نحصل على القيمة الاتجاهية لشهر فبراير ١٩٥٩ وهي ٤٦٦,٥ (٤٦٤,٣ + ٢,٢)، ولشهر مارس ١٩٥٩ وهي ٤٦٨,٧ (٤٦٦,٥ + ٢,٢) ولشهر إبريل ٤٧٠,٩ ومايو ٤٧٣,١ ويونيو ٤٧٥,٣ ويوليو ٤٧٧,٥ وأغسطس ٤٧٩,٧ وسبتمبر ٤٨١,٩ وأكتوبر ٤٨٤,١ ونوفمبر ٤٨٦,٣ وديسمبر ٤٨٨,٥. والقيم الاتجاهية السابقة يمكن اعتبارها تقريراً لمتوسط الطاقة الكهربائية لشهور سنة ١٩٥٩ إذا لم يكن هناك تغيرات موسمية. ولكن في دراستنا التغيرات الموسمية استنتجنا أن الدليل الموسمي الذي حصلنا عليه باستخدام طريقة النسب المئوية للاتجاه العام على أساس الوسيط - لشهر يناير مثلاً هو ١١٩,٦٪ لو لم يكن هناك تغيرات موسمية وبافتراض أن الدليل الموسمي سيظل كما هو خلال عام ١٩٥٩، فإن التنبؤ بالمتوسط الشهري للطاقة الكهربائية يجب أن يعدل بأن نأخذ ١١٩,٦٪ من القيمة الاتجاهية المقدرة لشهر يناير ١٩٥٩، ١٠٦,٧٪ من القيمة الاتجاهية المقدرة لشهر فبراير ١٩٥٩.. وهكذا لباقي الشهور كما هو موضح بالجدول التالي:

جدول رقم (١٣-٢٠)

القيم الاتجاهية وأثر الموسمية ومقدار الطاقة الكهربائية المقدرة لسنة ١٩٥٩

الشهر	القيمة الاتجاهية	أثر الموسمية	الطاقة المقدرة ١٩٥٩	الشهر	القيمة الاتجاهية	أثر الموسمية	الطاقة المقدرة ١٩٥٩
يناير	٤٦٤,٣	١,١٩٦	٥٥٥,٣	يوليو	٤٧٧,٥	,٨٣٥	٣٩٨,٧
فبراير	٤٦٦,٥	١,٠٦٧	٤٩٧,٧	أغسطس	٤٧٩,٧	,٨٩٥	٤٢٩,٢
مارس	٤٦٨,٨	١,٠٣٤	٤٨٤,٦	سبتمبر	٤٨١,٩	,٩٧١	٤٦٧,٩
إبريل	٤٧٠,٩	,٩٣٣	٤٣٩,٣	أكتوبر	٤٨٤,١	١,٠١٣	٥١٤,٦
مايو	٤٧٣,١	,٨٦٨	٤١٠,٦	نوفمبر	٤٨٦,٣	١,١٣٠	٥٤٩,٥
يونيو	٤٧٥,٣	,٨١٥	٣٨٧,٤	ديسمبر	٤٨٨,٥	١,٩٨	٥٨٢,٨

ويمكن أخيراً تخلص بيانات الظاهرة من أثر التغيرات الموسمية بنفس طريقة تخلص بيانات الظاهرة من أثر الاتجاه العام وذلك بقسمة كل عنصر في البيانات الأصلية (جدول رقم ١٣-١٣) على أرقام الدليل الموسمي (أو الأرقام القياسية الموسمية) المقابلة. وتسمى البيانات التي نحصل عليها بيانات لا موسمية أو بيانات معدلة لاستبعاد التغيرات الموسمية. مثل هذه البيانات تتضمن أثر الاتجاه العام والتغيرات الدورية والتغيرات العشوائية، أي أن:

$$\text{القيمة الأصلية (ص)} \div \text{التغيرات الموسمية (و)} = \text{القيمة الاتجاهية (ص)} \times \text{التغيرات الدورية (د)}$$

$$\times \text{التغيرات العشوائية (ش)}$$

فمثلاً إذا أردنا معرفة متوسط الطاقة الكهربائية المستهلكة (مليون كيلو واط

ساعة) لشهر يناير ١٩٥١ بعد استبعاد التأثيرات الموسمية فإن هذا المتوسط يكون على النحو التالي:

$$= \frac{100 \times 318}{119,6} = 265,9 \text{ مليون كيلوات ساعة}$$

وبالمثل لباقي قيم الظاهرة في شهور السنوات المختلفة للسلسلة الزمنية من ١٩٥١ - ١٩٥٨.

رابعاً: تقدير التغيرات العشوائية:

بعد أن تمكنا من استبعاد أثر الاتجاه العام والتغيرات الدورية والتغيرات الموسمية من السلسلة الزمنية، فإنها سوف تقع تحت تأثير التغيرات العشوائية أو التغيرات الفجائية (مثل وقوع الزلازل والبراكين والحروب) التي من الصعب التكن بها أو التنبؤ بوقوعها، وبالتالي يصعب تحديد حجم هذه التغيرات وتحديد مدى تأثيرها في قيمة الظاهرة. ويمكن تقدير قوة تأثير التغيرات العشوائية نظرياً حيث أننا سوف لانحتاج إلا إلى مقارنة القيم الرصالية بالقيم النظرية المحسوبة على أساس خط الاتجاه العام والتغيرات الموسمية، فأى فرق أو إنحراف بين القيمة الأصلية والقيمة النظرية تنسبه إلى التغيرات العشوائية. كما يمكن حساب التغيرات العشوائية بغرض استبعادها وحذفها من قيم الظاهرة وذلك بقسمة البيانات الأصلية (ص) للظاهرة على حاصل ضرب كل من أثر الاتجاه العام والتغيرات الدورية فقط في حالة إذا كانت السلسلة الزمنية موضع التحليل سنوية، وعلى حاصل ضرب كل من أثر الاتجاه العام والتغيرات الدورية والموسمية إذا كانت السلسلة الزمنية موسمية سنوية، أى أن:

(١) فى حالة السلسلة الزمنية السنوية:

$$\frac{ص}{ت \times د} = \text{التغيرات العشوائية (ش)}$$

(٢) فى حالة السلسلة الزمنية الموسمية:

$$\frac{\text{ص}}{\text{ت} \times \text{د} \times \text{و}} = \text{التغيرات العشوائية (ش)}$$

ومن الناحية العملية وجد أن التغيرات العشوائية (غير المنتظمة) تتجه إلى أن تكون ذات تأثير قليل (حجم صغير)، وأنها غالباً تتجه إلى أن تتبع نمط التوزيع المعتدل (الطبيعي)، أى أنها عبارة عن انحرافات صغيرة تحدث بتكرارات كبيرة أما الانحرافات الكبيرة فتحدث بتكرارات صغيرة.

المحتويات

فهرس محتويات الكتاب

صفحة

٧

تصدير

١٥

مقدمة: فى المفاهيم الاحصائية

الوصف الاحصائى، الاستدلال، (الاستنتاج) الاحصائى،
اختبار الفروض الاحصائية، التنبؤ (التوقع) الاحصائى،
المجتمع والعينة، البيانات (المعطيات) الاحصائية، المفردات
والمتغيرات، أنواع البيانات، الدقة والأخطاء فى البيانات

الباب الأول

جمع البيانات وطرق إعدادها للتحليل الكمى

٣١

مقدمة

٣٣

الفصل الأول: جمع البيانات

مصادر جمع البيانات، طرق (أدوات) جمع البيانات،
الاستمارات الاحصائية، تصميم الاستمارة الاحصائية

١٠٣

الفصل الثانى: تصنيف وجدولة البيانات

تجميع البيانات، تصنيف البيانات، الجدولة اليدوية للبيانات،
الجدولة الآلية للبيانات الاحصائية

١٢١

الفصل الثالث: العرض البيانى للبيانات الإحصائية

العرض البيانى للبيانات الخام (غير المبوبة)، العرض البيانى
للبيانات المبوبة

الباب الثانى

مقاييس الوصف

١٧٣

مقدمة

صفحة

١٧٥

الفصل الرابع: مقاييس النزعة المركزية

أولاً: المتوسط: المتوسط الحسابي، المتوسط الهندسي، المتوسط التوافقي، ثانياً: الوسيط: شبيهات الوسيط، ثالثاً: المنوال، العلاقة بين مقاييس النزعة المركزية، مزايا ومثالب مقاييس النزعة المركزية.

٢٤١

الفصل الخامس: مقاييس الانحراف (التشتت) والاختلاف

أنواع الانحراف، الانحراف الربيعي، الانحراف المتوسط، التباين والانحراف المعياري، مؤشرات الاختلاف (التباين)، القيم (الوحدات)، المعيارية.

٢٨٥

الفصل السادس: مؤشرات التركيز (الالتواء والتفرطح)

الالتواء، مقاييس الالتواء، التفرطح، العزوم وقياس الالتواء والتفرطح.

الباب الثالث

التقدير الاحصائي وأساليب المقارنة

٣٠٧

مقدمة

٣٠٩

الفصل السابع: تقدير خصائص (معالم) المجتمع

أنواع التقدير تقدير فترة الثقة لمعالم المجتمع، التقدير من احصائية (مقاييس) العينات.

٣٣١

الفصل الثامن: اختبارات الفروض الاحصائية

قواعد اختبار الفروض الاحصائية، اختبار المعنوية (الدلالة) تحديد التوزيع النظري (الاحتمالي) للاحصائية المختبرة، تحديد المنطقة الحرجة، اختبار انتماء عينة لمجتمع متوسطة معلوم، اختبار الاختبارات الاحصائية.

٣٥٣

الفصل التاسع: أساليب المقارنة البارامترية (المعلمية)

اختبار استيودنت - t، اختبار الفرق بين المتوسطات، تحليل التباين (اختبار ف).

صفحة

٣٨١

الفصل العاشر: أساليب المقارنة اللابارامترية (اللامعملية)

أولاً: اختبار مربع كاي ثانياً: اختبار كولموجوروف -
سميرنوف (اختبار «د»)، ثالثاً: اختبار مان - هويتنى
(اختبار «ى»)، رابعاً: اختبار ويلكوكسون (اختبار «ق»)،
خامساً: اختبار كروسكال - واليس (اختبار «ه»).

الباب الرابع

أساليب قياس العلاقات والتغيرات

٤٢٥

مقدمة

٤٢٧

الفصل الحادى عشر: تحليل الارتباط

مقاييس الارتباط، حساب معامل الارتباط، معامل ارتباط
ضرب العزوم، معامل ارتباط الرتب: معامل ارتباط سبيرمان،
معامل كندال لإرتباط الرتب، اختبار المعنوية الاحصائية
للارتباط.

٤٦٧

الفصل الثانى عشر: تحليل الانحدار

أنواع تحليل الانحدار، تحليل الانحدار البسيط

٤٨٣

الفصل الثانى الثالث عشر: تحليل السلاسل الزمنية

التمثيل البياني للسلاسل الزمنية، التغيرات (التحركات)
المميزة فى السلاسل الزمنية، تحليل السلسلة الزمنية: أولاً:
تقدير الاتجاه العام، ثانياً: تقدير التغيرات الدورية، ثالثاً:
التغيرات الموسمية، رابعاً: تقدير التغيرات العشوائية.

٥٣٥

فهرس محتويات الكتاب

٥٤١

فهرس الأشكال

٥٤٥

المراجع الرئيسية

ملحق الجداول الاحصائية

فهرس الأشكال

شكل رقم	صفحة
١-١	أنواع إطار المعاينة الجغرافية ٥٧
٢-١	أنواع الاختبار العشوائي للتوزيعات المكانية ٦٧
٣-١	تحديد مواقع ٢٠ قرية باستخدام الأحداثيات العشوائية ٧٣
٤-١	التوزيع المكاني لمفردات العينة العشوائية ٧٦
١-٣	تطور إنتاج الثروة السمكية في كندا في الفترة من ١٩٣٨-٢٧ ١٢٣
٢-٣	ورق بياني لوغاريتمي مزدوج ١٢٥
٣-٣	معدلات المواليد والوفيات ١٩٦١ - ١٩٧٠ ١٢٦
٤-٣	معدلات المواليد والوفيات ٦١ - ١٩٧٠ (رسم بياني لوغاريتمي) ١٢٧
٥-٣	أشكال الانتشار لتوضيح اتجاهات الترابط بين المتغيرات ١٢٩
٦-٣	قيمة منتجات مصايد الأسماك الكندية في الفترة ١٩١٧ - ١٩٣٨ ١٣١
٧-٣	قيمة الصادرات المصرية في عام ١٩٦٣ / ٦٢ ١٣٢
٨-٣	أوزان المجموعات الرئيسية للواردات المصرية ١٩٦٣ / ٦٢ ١٣٣
٩-٣	الرسوم البيانية الحجمية (كرات) والمساحية (دوائر) ١٣٥
٩-٣	الكرات النسبية لبيانات سكان المدن المصرية لعام ١٩٦٦ ١٣٥
١٠-٣	الرسوم البيانية الدائرية لتمثيل عنصرى درجة الحرارة والمطر ١٣٧
١١-٣	إنتاج النحاس في العام بآلاف الأطنان المترية ١٣٩
١٢-٣	النسبة المئوية للمشتغلين بكل من الحرف الرئيسية في إحدى المحافظات ١٣٩
١٣-٣	الواردات البريطانية من الأقطان الخام ١٤١

شكل رقم	صفحة
١٤-٣	التطور النسبي لمكونات الدخل القومي لمصر ١٤١
١٥-٣	أشكال الاهرامات السكانية البسيطة ١٤٤
١٦-٣	الهرم البياني المركب ١٤٥
١٧-٣	الهرم السكاني المنقطع ١٤٦
١٨-٣	إنتاج مناطق الصيد في مصر ٦١ - ١٩٦٤ ١٥٠
١٩-٣	تحديد فئات المصانع ونسبة العمالة بها ١٥٤
٢٠-٣	بعض أنواع الرموز التصويرية ١٥٤
٢١-٣	نموذج للرسوم البيانية التصويرية ١٥٥
٢٢-٣	أعداد قراء الصحف اليومية «طريقة الرسوم التصويرية» ١٥٦
٢٣-٣	المدرج التكرارى لأطوال ١٠٠ رافد نهري (بالمتر) ١٥٨
٢٤-٣	المضلع التكرارى لأطوال ١٠٠ رافد نهري (بالمتر) ١٦٠
٢٥-٣	المنحنى التكرارى لأطوال ١٠٠ رافد نهري (بالمتر) ١٦١
٢٦-٣	أشكال المنحنيات التكرارية ١٦٣
٢٧-٣	المنحنى المتجمع الصاعد لعدد ١٠٠ رافد نهري بالمتر ١٦٥
٢٨-٣	المنحنى المتجمع الهابط لعدد ١٠٠ رافد نهري بالمتر ١٦٦
١-٤	تعيين الوسيط بيانياً من المدرج التكرارى والمنحنى المتجمع النسبى الصاعد ٢٠٥
٢-٤	تعيين الربيع الأدنى والأعلى بيانياً من المنحنى المتجمع النسبى الصاعد ٢٠٩
٣-٤	تحديد موضع المنوال داخل الفئة المنوالية «بطريقة الرافعة» ٢١٢
٤-٤	طريقة تعيين المنوال بيانياً من المدرج التكرارى ٢١٥
٥-٤	مقاييس النزعة المركزية وعلاقتها بأنواع التوزيعات التكرارية ٢١٨
٦-٤	العلاقة بين مقاييس النزعة المركزية الثلاثة فى التوزيعات التكرارية الموجبة الالتواء ٢٢١

شكل رقم	صفحة
٧-٤	المدرج التكرارى ومنحنى التوزيع التكرارى لكمية الأمطار السنوية فى مرض ييدستون - إنجلترا ١٩٠١ - ١٩٣٠
٨-٤	المدرجات التكرارية للإنتاج السنوى لخام الحديد فى كل من بلجيكا، فرنسا، لوكسمبورج والمملكة المتحدة ٣٨ - ١٩٥٧
١-٥	الوسيط والربيعين (الأدنى والأعلى) لكميات المطر السنوية
٢-٥	خطوات حساب مقاييس التشتت (الانحراف المتوسط، التباين، الانحراف المعيارى)
٣-٥	مقارنة بين مقياس التشتت: الانحراف المتوسط والتباين.
١-٦	أشكال المدرجات التكرارية وأنواع التوزيعات التكرارية التى تمثلها
٢-٦	أنواع التفرطح لمنحنيات التوزيعات التكرارية
١-٨	التوزيع العينى للاختلاف بين المتوسطات مقداراً بالقيم المعيارية
٢-٨	المنطقة الحرجة واختبارات الفروض الاحصائية
١-٩	طريقة جمع البيانات للتحليل الأحادى للتباين
١-١١	التوزيع المعتدل للمتغيرين
٢-١١	قوة الارتباط بين المتغيرين r ، s ، كما يوضحها شكل انتشار المفردات لكل منهما
١-١٢	انحراف نقط تمثيل المتغيرين s ، v عن خط الانحدار
٢-١٢	شكل الانتشار وخط الانحدار للعلاقة بين مساحة حوض النهر وطوله
١-١٣	المنحنى التاريخى لسلسلة أعداد الماشية فى الولايات المتحدة
٢-١٣	الأرقام القياسية لتطور أعداد الماشية والأغنام فى اسكتلندا
٣-١٣	تمثيل النمو السكانى لمدينتين أ، ب على الرسم البيانى النصف لوغاريتمى

شكل رقم	صفحة
٤-١٣	٤٩٦
٥-١٣	٤٩٨
٦-١٣	٥٠٢
٧-١٣	٥١٨
٨-١٣	٥٢٦

شكل رقم	صفحة
٧-٤	المدرج التكرارى ومنحنى التوزيع التكرارى لكمبة الأمطار السنوية فى مرض بيدستون - إنجلترا ١٩٠١ - ١٩٣٠
٨-٤	المدرجات التكرارية للإنتاج السنوى لخام الحديد فى كل من بلجيكا، فرنسا، لوكسمبورج والمملكة المتحدة ٣٨ - ١٩٥٧
١-٥	الوسيط والربيعين (الأدنى والأعلى) لكميات المطر السنوية
٢-٥	خطوات حساب مقاييس التشتت (الانحراف المتوسط، التباين، الانحراف المعيارى)
٣-٥	مقارنة بين مقاييس التشتت: الانحراف المتوسط والتباين
١-٦	أشكال المدرجات التكرارية وأنواع التوزيعات التكرارية التى تمثلها
٢-٦	أنواع التفرطح لمنحنيات التوزيعات التكرارية
١-٨	التوزيع العينى للاختلاف بين المتوسطات مقداراً بالقيم المعيارية
٢-٨	المنطقة الحرجة واختبارات الفروض الاحصائية
١-٩	طريقة جمع البيانات للتحليل الأحادى للتباين
١-١١	التوزيع المعتدل للمتغيرين
٢-١١	قوة الارتباط بين المتغيرين s_1 ، s_2 كما يوضحها شكل انتشار المفردات لكل منهما
١-١٢	انحراف نقط تمثيل المتغيرين s ، v عن خط الانحدار
٢-١٢	شكل الانتشار وخط الانحدار للعلاقة بين مساحة حوض النهر وطوله
١-١٣	المنحنى التاريخى لسلسلة أعداد الماشية فى الولايات المتحدة
٢-١٣	الأرقام القياسية لتطور أعداد الماشية والأغنام فى اسكتلندا
٣-١٣	تمثيل النمو السكانى لمدينتين أ، ب على الرسم البيانى النصف لوغاريتمى

صفحة	شكل رقم
٤٩٦	٤-١٣ مكونات السلسلة الزمنية المثالية
	٥-١٣ الاتجاه العام لتطور إجمالي الصادرات لجمهورية مصر ٦٠ - ١٩٨٠
٤٩٨	٦-١٣ المنحنى التاريخي وخط الاتجاه العام بطريقة أنصاف المتوسطات
٥٠٢	للإنتاج السنوي من الفحم في دولة ما ٤٨ - ١٩٥٨
٥١٨	٧-١٣ تحديد فترة الدورة
٥٢٦	٨-١٣ تطور استخدام الطاقة الكهربائية (مليون كيلو واط ساعة) في دولة ما في الفترة من ١٩٥١ - ١٩٥٨

المراجع الرئيسية

أولاً: المراجع العربية:

- أحمد عباده سرحان (١٩٦٨): أسس الإحصاء، دار الكتب الجامعية، القاهرة.
- أحمد عباده سرحان (١٩٧١): طرق التحليل الإحصائي، الهيئة العامة للكتاب، القاهرة.
- أحمد عباده سرحان، صلاح الدين طلبه، فاروق عبد العظيم أحمد (بدون تاريخ): الإحصاء، مؤسسة شباب الجامعة، الإسكندرية.
- أحمد عباده سرحان وآخرون (١٩٦٨): تحليل الانحدار والارتباط، مكتبة عين شمس، القاهرة.
- اسماعيل محمد هاشم (١٩٦٨): الإحصاءات التطبيقية، دار المعارف، القاهرة.
- السيد سعد قاسم ولطفى هندی (١٩٦٧): مبادئ الإحصاء التجريبي، الطبعة الثانية، دار المعارف، القاهرة.
- السيد محمد خيرى (١٩٧٠): الإحصاء فى البحوث النفسية والتربوية والاجتماعية، الطبعة الرابعة، دار النهضة العربية، القاهرة.
- أنيس رمسيس منصور وزكى محمد عبد الرحمن (١٩٦٧): مقدمة إلى الإحصاء، مكتبة الانجلو المصرية، القاهرة.
- بدر الدين المصرى (١٩٧٠): مذكرات فى الإحصاء، دار الجامعات المصرية، الإسكندرية.
- عبد الإله أبو عياش (١٩٧٨): الإحصاء والكمبيوتر فى معالجة البيانات - مع تطبيقات جغرافية - وكالة المطبوعات، الكويت.
- عبد الرحمن البدرى (١٩٦١): مبادئ الطرق الإحصائية، دار النهضة العربية، القاهرة.
- عبد المجيد فراج (١٩٧٠): الأسلوب الإحصائي، الطبعة الثانية، مكتبة القاهرة الحديثة، القاهرة.
- محمد الفراء (١٩٧٥): مناهج البحث فى الجغرافيا بالوسائل الكمية، الطبعة الثانية، وكالة المطبوعات، الكويت.
- محمد على بشر ومحمد الروبى (١٩٨٠): مقدمة فى طرق الإحصاء وتصميم التجارب، دار المطبوعات الجديدة، الإسكندرية.
- محمد مظلوم حمدى (١٩٦١): طرق الإحصاء، دار المعارف، القاهرة.
- ناصر عبد الله الصالح ومحمد محمود السريانى (١٩٧٩): الجغرافية الكمية والاحصائية - أسس وتطبيقات - مكة المكرمة، المملكة العربية السعودية.

ثانياً: المراجع الأجنبية:

- Alder, H.L. & Roessler, E.B. (1964): Introduction to Probability and Statistics, W.H. Freeman.
- Berry, B. & Baker, A. (1968): "Geographic Sampling" in Berry & Marble, Spatial Analysis, Englewood Cliffs, Prentice - Hall Inc., N.J., pp. 91 - 160.
- Burton Ian (1963): "The Quantitative Revolution and Theoretical Geography" The Canadian Geography.
- Cole, J. P. & King, C.A.M. (1970): Quantitative Geography Techniques and Theories in Geography, 3rd Ed., Wiley, London.
- Davis, J.C. (1973): Statistics and Data Analysis in Geology, Wiley, New York.
- Davis, W.K. (1972): The Conceptual Revolution in Geography, Univ. of London Press, London.
- Dornkawp, J. C. & King, C.A.M. (1971): Numerical Analysis in Geomorphology, Arnold, London.
- Ebdon, D. (1977): Statistics in Geography, Basil, Blackwell, Oxford.
- Gerogry, S. (1970); Statistical Methods and Geographers, Longman, London.
- Hammond, R. & Mc Cullagh, P.S. (1974) Quantitative Techniques in Geography, Clarendon Press, Oxford.
- Harper, W. M. (1974): Statistics, 4th ed., M. & E. Handbook, London.
- Harvey, D. W. (1969): An Outline of Statistics, Longman, London.
- Kendall, M.G. (1970) Rank Correlation Methods, Griffin, London.
- King, L. J. (1967): Statistical Analysis in Geography, Englewood Cliffs, N. J.
- Koch, G.S. & Link, R.F. (1971): Statistical Analysis of Geological Data, vol. II, Wiley, New York.

- Lindly, D.V. & Miller, J.C.P. (1953): Cambridge Elementary Statistical Tables, Cambridge.
- Morkhouse, F.J. & Wilkinson, H. R., (1971): Maps and Diagrams, Methuen, London..
- Moroney, M.J. (1975): Facts from Figures, Penguin Book, London.
- Siegle, S., (1956): Non-parametric Statistical for the Behavioural Sciences, Mc Graw-Hill, New York.
- Spiegle, M.R. (1972): Theory and Problems of Statistics, McGraw-Hill, New York.
- Stoodley, K.D.C. (1974): Basic Statistical Techniques, Bradford Univ. Press,
- Till, R. (1974): Statistical Methods for Earth Scientists, MacMillan Press, Ltd., London.
- Toyne, P. & Newby, P.T., (1971) Techniques in Human Geography MacMillan Press, Ltd., London.
- Von Mises, R. (1964): Mathematical Theory of Probability and Statistics, Academic Press, New York.
- Yeates, Maurice (1974): An Introduction to Quantitative Analysis in Human Geography, McGraw-Hill, New York.
- Yevjevich, V.L. (1971): Probability and Statistics in Hydrology, Water Resources publication, Colorado.
- Yule, G.U. & Kendaly, M.G. (1953): An Introduction to the Theory of Statistics, Griffin, London.

الجدول

جدول رقم (٣-٢)

في شرح جداول اللوغاريتمات وكيفية استعمالها

٧ جداول اللوغاريتمات العادية محسوبة على مقتضى الأساس ١٠ وأول من حسب هذه اللوغاريتمات هنري بريجز (Henry Briggs) سنة ١٦١٥ ميلادية بناء على ترقية نيبير (Napier) له ويقال للوغاريتمات المحسوبة على هذا الأساس اللوغاريتمات العادية أو اللوغاريتمات البرمجزية (نسبة إلى الرجل بريجز الذي أدخلها).

٨ في هذه الجداول تكون لوغاريتمات الأعداد التي هي قوى لعدد ١٠ أعداداً صحيحة مثلاً

لو ١٠٠٠ = ٣ لأن ١٠ ^٣ = ١٠٠٠	لو ١ = ٠ لأن ١٠ ^٠ = ١
لو ١٠٠ = ٢ لأن ١٠ ^٢ = ١٠٠	لو ١٠ = ١ لأن ١٠ ^١ = ١٠
لو ١٠ = ١ لأن ١٠ ^١ = ١٠	لو ١ = ٠ لأن ١٠ ^٠ = ١
لو ١ = ٠ لأن ١٠ ^٠ = ١	لو ١ = ٠ لأن ١٠ ^٠ = ١

٩ لوغاريتمات الأعداد التي ليست قوى لعدد ١٠ تتركب من عدد صحيح ومن كسر عشري ويقال للعدد الصحيح العدد البياني وللکسر الجزء العشري.

١٠ العدد البياني من لوغاريتم أي عدد أكبر من الواحد يكون موجباً ويساوي عدد أرقامه الصحيحة ناقصاً واحداً.

٦٣٤٥٠ هو ٤	الجزء البياني من لوغاريتم
٦٣٤,٥ هو ٢	١ ١ ١
٦,٣٤٥ هو ٠	١ ١ ١

١١ العدد البياني من لوغاريتم أي عدد أصغر من الواحد يكون سالباً ويساوي عدد الأصغر إلى ثل الشرطة المشرية مباشرة مضافاً إليه واحد.

٠,٦٣٤٥ هو ١	الجزء البياني من لوغاريتم
٠,٠٠٦٣٤٥ هو ٣	١ ١ ١
٠,٠٠٠٠٦٣٤٥ هو ٥	١ ١ ١

(٢) تليه) عند ما يكون العدد البياني سالباً تكتب العلامة (-) فوق العدد البياني مثل ٢-

١٢ الأعداد المركبة من أرقام متحدة ذات ترتيب واحد بلا تختلف إلا موضع العلامة العشرية تكون لوغاريتها متحدة في الجزء العشري وعختلفة في العدد اليائي .

فالأجزاء العشرية من لوغاريثات الأعداد ١٠٠٦٣٤٥ ، ١٠٦٣٤٥ ، ١٠٦٣٤٥٠ كلها متساوية

١٣ لإيجاد الجزء العشري من لوغاريتم عدد لا تزيد أرقامه المئوية على رقم واحد مثل

٨٠٠ أو ٨٠٠٠

الرقم	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
٩٨٧٦٥٤٣٢١	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
٥٤٤٣٣٢٢١١	٨٠	٩٠	٩١	٩٢	٩٣	٩٤	٩٥	٩٦	٩٧	٩٨

نبحث عن العدد ٨٠٠ في صفحات جدول اللوغاريثات الأعداد في الصف الرأسي الأول ويبحث عن (٠) في الصف الأفقي الأول من هذه الصفحة ثم نتبع الصف الأفقي المبدوء بالعدد ٨٠ والصف الرأسي المبدوء بصفر فنجد في تقاطع هذين الصفين العدد ٩٠٣١ فيكون هو الجزء العشري من لوغاريتم ٨٠٠ أو ٨٠٠٠

١٤ لإيجاد الجزء العشري من لوغاريتم عدد مركب من رقمين متتبعين مثل ٨٥ أو ٨٥٠

أو ٨٠٥

الرقم	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
٩٨٧٦٥٤٣٢١	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
٥٤٤٣٣٢٢١١	٨٥	٩٢	٩٣	٩٤	٩٥	٩٦	٩٧	٩٨	٩٩	١٠٠

نبحث عن العدد ٨٥ في صفحات جدول اللوغاريثات في الصف الرأسي الأول ويبحث عن (٠) في الصف الأفقي الأول من هذه الصفحة ثم نتبع الصف الأفقي المبدوء بالعدد ٨٥ والصف الرأسي المبدوء بصفر فنجد في تقاطع هذين الصفين ٩٢٩٤ فيكون هو الجزء العشري من لوغاريتم ٨٥ أو ٨٥٠ أو ٨٠٥

١٥ لإيجاد الجزء العشري من لوغاريتم عدد مركب من ثلاثة أرقام متتبعين مثل ٨٥٦

أو ٨٠٥٦ أو ٨٥٦٠

نبحث عن العدد المركب من الرقمين الأولين من يسار العدد (٨٥) في صفحات الجدول في الصف الرأسي الأول ويبحث عن الرقم الثالث ٦ في الصف الأفقي الأول من هذه الصفحة ثم نتبع الصف الأفقي المبدوء بالعدد ٨٥ والصف الرأسي المبدوء برقم ٦ فنجد في تقاطع هذين الصفين ٩٣٢٥ فيكون هو الجزء العشري من لوغاريتم ٨٥٦ أو ٨٠٥٦ أو ٨٥٦٠

١٦ لإيجاد الجزء العشري من لوغاريتم عدد مركب من أربعة أرقام متتبعين مثل ٨٥٦٢

أو ٨٥٦٢ أو ٨٥٦٢٠

نبحث عن الجزء العشري للعدد ٨٥٦ بالطريقة نفسها فنجد أنه ٩٣٢٥ ثم نبحث عن ٢ في الصف الأفقي الأول من أعمدة القروق ونتبع الصف الأفقي المبلو بالعدد ٨٥ والصف الرأسى للملء بالفرق ٢ فنجد في مقاطع هذين الصفيين ١ فيكون هو العدد الذى يلزم إضافته إلى ٩٣٢٥ لينتج الجزء العشري من لوغاريتم العدد ٨٥٦٢ .

وعلى ذلك يكون ٩٣٢٥ + ١ (٠,٩٣٢٦) هو الجزء العشري من لوغاريتم العدد ٨٥٦٢ أو ٨٥,٦٢ أو ٨٥٦٢٠ .

١٧ بما تقدم نعلم طريقة إيجاد لوغاريتم أى عدد لا يزيد على أربعة أرقام وذلك بأن نأخذ أولاً الجزء العشري ثم نضيف إليه عدده البياضى .

٤,٨٠٢٤ =	فلا لو ٦٣٤٥٠
٠,٨٠٢٤ =	لو ١,٣٤٥
١,٨٠٢٤ =	لو ٠,٦٣٤٥
٣,٨٠٢٤ =	لو ٠,٠٦٣٤٥

١٨ لإيجاد العدد المقابل للوغاريتم معلوم :

نبحث بطريقة مماثلة للطريقة المينة بيند ١٦ عن العدد المقابل للجزء العشري من هذا اللوغاريتم في جدول الأعداد المقابلة للوغاريتمات ثم نعدل هذا العدد كما يقتضيه العدد البياضى للوغاريتم بأن نضع على يمينه أصفراً أو بفصل منه أرقاماً عشرية .
فلذا أردنا إيجاد العدد الذى لوغاريتمه هو ٢,٠٦٧٤ نجري العمل هكذا .

الفرق	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
١٨٧٦٥٤٣٢١	١١٥٦١١٥٣١١٥٩	١١٦٤١١٦١١٥٩	١١٦٤١١٦١١٥٩	١١٦٤١١٦١١٥٩	١١٦٤١١٦١١٥٩	١١٦٤١١٦١١٥٩	١١٦٤١١٦١١٥٩	١١٦٤١١٦١١٥٩	١١٦٤١١٦١١٥٩	١١٦٤١١٦١١٥٩
٢٢٢٢١١١٠	١١٢٨٠,٨٨									

نبحث عن ٠,٠٦ في صفحات جدول الأعداد المقابل للوغاريتمات في الصف الرأسى الأول ونبحث عن الرقم العشري الثالث ٧ في الصف الأفقى الأول من هذه الصفحة ثم نتبع الصف الأفقى المبلو بالعدد ٠,٠٦ والصف الرأسى المبلو بالعدد ٧ فنجد في مقاطع هذين الصفيين ١١٦٧ ثم نبحث عن الرقم العشري الرابع ٤ في الصف الأفقى الأول من أعمدة القروق ونتبع الصف الأفقى المبلو بالعدد ٠,٠٦ والصف الرأسى المبلو بالفرق ٤ فنجد في مقاطع هذين الصفيين ١ فيكون هو العدد الذى يلزم إضافته إلى ١١٦٧ لنتج الأرقام المكونة للعدد المطلوب .
وعلى ذلك تكون أرقام العدد البحارى البحث عنه تسارى ١١٦٧ + ١ (أى ١١٦٨) .

وبما أن العدد البياضى للوغاريتم هو ٢ يكون عدد أرقامه الصحيحة ٣ ويكون العدد الذى لوغاريتمه ٢,٠٦٧٤ هو ١١٦,٨٠ .

הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
הערות	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59																																									

(تابع) جدول لوظائفیات الأعداد

[illegible]

[illegible]

(تابع) جدول الأعداد المتعاقبة الهندسية

[illegible]

جدول (م- ٨) : توزيع ستیوننت - ت

مستويات الشقة						درجات المحطة
١٠٠	٢٠٥	١٠٥	٢٠	٤٠	١٠٥	
٦٢,٦٦	١٢,٧١	٦,٢١	٢,٠٨	١,٢٨	١,٠٠	١
٦,١٢	٢,٣٠	١,١٢	١,٨٩	١,٠٦	٨٢	٢
٢,٨٤	٢,١٨	٢,٣٥	١,٦٤	١,١٨	٧٦	٣
٢,٦٠	٢,٧٨	٢,١٢	١,٥٢	١,١٤	٧٤	٤
٢,٥٢	٢,٥٧	٢,٠٦	١,٤٨	١,١٢	٧٢	٥
٢,٧١	٢,٤٥	١,١٤	١,٤٤	١,١١	٧٢	٦
٢,٥٠	٢,٢٦	١,٨١	١,٤٢	١,٠	٧١	٧
٢,٢٦	٢,٢١	١,٨٦	١,١٠	٨١	٧١	٨
٢,٢٥	٢,١٦	١,٨٢	١,٤٨	٨٨	٧٠	٩
٢,١٧	٢,١٢	١,٨١	١,٢٧	٨٨	٧٠	١٠
٢,١١	٢,٢٠	١,٨٠	١,٢٦	٨٨	٧٠	١١
٢,٠٦	٢,١٨	١,٧٨	١,٢٦	٨٧	٦٩	١٢
٢,٠١	٢,١٦	١,٧٧	١,٢٥	٨٧	٦٩	١٣
٢٠١٨	٢,١٤	١,٧٦	١,٢٤	٨٧	٦٩	١٤
٢,١٥	٢,١٢	١,٧٥	١,٢٤	٨٧	٦٩	١٥
٢,١٢	٢,١٤	١,٧٥	١,٢٤	٨٧	٦٩	١٦
٢,١٠	٢,١٢	١,٧٤	١,٢٣	٨٦	٦٩	١٧
٢,٨٨	٢,١٠	١,٧٣	١,٢٣	٨٦	٦٩	١٨
٢,٨٦	٢,٠٩	١,٧٣	١,٢٣	٨٦	٦٩	١٩
٢,٨٤	٢,٠٩	١,٧٢	١,٢٢	٨٦	٦٩	٢٠
٢,٨٢	٢,٠٧	١,٧٢	١,٢٢	٨٦	٦٩	٢١
٢,٨٠	٢,٠٦	١,٧١	١,٢٢	٨٦	٦٨	٢٢
٢,٧٨	٢,٠٦	١,٧١	١,٢١	٨٦	٦٨	٢٣
٢,٧٦	٢,٠٥	١,٧٠	١,٢١	٨٥	٦٨	٢٤
٢,٧٥	٢,٠٤	١,٧٠	١,٢١	٨٥	٦٨	٢٥
٢,٧٢	٢,٠٣	١,٦٩	١,٢١	٨٥	٦٨	٢٦
٢,٧٠	٢,٠٢	١,٦٨	١,٢٠	٨٥	٦٨	٢٧
٢,٦٦	٢,٠٠	١,٦٧	١,٢٠	٨٥	٦٨	٢٨
٢,٦٤	١,٩٩	١,٦٧	١,٢١	٨٥	٦٨	٢٩
٢,٦٢	١,٩٨	١,٦٦	١,٢١	٨٥	٦٨	٣٠
٢,٥٧	١,٩٦	١,٦٤	١,٢٨	٨٤	٦٧	٣١

١ - اعتبار شئى الطرف

ب - اعتبار أحدى الطرف

072

074

07A

جدول رقم (م- ١٠): جدول مربع كاي (X²)

استعمال الحسول على قيمة أهمل من كاي ²										درجات الحرية
٠.٠١	٠.٠٢٥	٠.٠٥	٠.١٠٠	٠.٢٥٠	٠.٥٠٠	٠.٧٥٠	٠.٩٠٠	٠.٩٧٥	٠.٩٩٠	
٦,٦٢	٥,٠٢	٢,٨٤	٢,٧١	١,٣٢	٠,٤٥	٠,١٠٢	٠,٠١٥٨	٠,٠٠٤٠	٠,٠٠٠١	١
٩,٢١	٧,٢٨	٥,٩٩	٤,٦١	٢,٧٧	١,٣٩	٠,٥٧٥	٠,٢١١	٠,٠٥٠٦	٠,٠١٢٠	٢
١١,٣	٩,٣٥	٧,٨١	٦,٢٥	٤,١١	٢,٢٧	١,٢١	٠,٥٨٤	٠,٢١٩	٠,٠١٥٠	٣
١٣,٢	١١,١	٩,٤٩	٧,٧٨	٥,٢٩	٣,٣٦	١,٩٢	١,٠٦	٠,٤٨٤	٠,٢٩٧	٤
١٥,١	١٢,٨	١١,١	٩,٢٤	٦,٣٣	٤,٣٥	٢,٦٧	١,٦٦	٠,٨٣١	٠,٥٥٤	٥
١٦,٥	١٤,٤	١٢,٦	١٠,٦	٧,٨٤	٥,٣٥	٣,٤٥	٢,٢٠	١,٢٤	٠,٨٧٢	٦
١٨,٥	١٦,٠	١٤,١	١٢,٠	٩,٠٤	٦,٣٥	٤,٢٥	٢,٨٣	١,٦٩	١,٢٤	٧
٢٠,٦	١٧,٥	١٥,٥	١٣,٤	١٠,٢	٧,٣٤	٥,٠٧	٣,٤٩	٢,١٨	١,٦٥	٨
٢١,٦	١٩,٠	١٦,٩	١٤,٧	١١,٤	٨,٣٤	٥,٩٠	٤,١٧	٢,٧٠	٢,٠٩	٩
٢٣,٢	٢٠,٥	١٨,٣	١٦,٠	١٢,٥	٩,٣٤	٦,٧٤	٤,٨٧	٢,٢٥	٢,٥٦	١٠
٢٤,٦	٢١,٩	١٩,٧	١٧,٣	١٣,٧	١٠,٣	٧,٥٨	٥,٥٨	٢,٨٢	٢,٠٥	١١
٢٦,١	٢٣,٣	٢١,٠	١٨,٥	١٤,٨	١١,٣	٨,٤٤	٦,٣٥	٤,٤٠	٢,٥٧	١٢
٢٧,٧	٢٤,٧	٢٢,٤	١٩,٨	١٦,٠	١٢,٣	٩,٣٠	٧,٠٤	٥,٠١	٤,١١	١٣
٢٩,١	٢٦,١	٢٣,٧	٢١,١	١٧,١	١٣,٣	١٠,٢	٧,٧٩	٥,٦٣	٤,٦٦	١٤
٣٠,٦	٢٧,٥	٢٥,٠	٢٢,٣	١٨,٢	١٤,٣	١١,٠	٨,٥٥	٦,٢٦	٥,٢٣	١٥
٣٢,٠	٢٨,٨	٢٦,٣	٢٣,٥	١٩,٤	١٥,٣	١١,٩	٩,٣١	٦,٩١	٥,٨١	١٦
٣٣,٤	٣٠,٢	٢٧,٦	٢٤,٨	٢٠,٥	١٦,٣	١٢,٨	١٠,١	٧,٥٦	٦,٤١	١٧
٣٤,٨	٣١,٥	٢٨,٩	٢٦,٠	٢١,٦	١٧,٣	١٣,٧	١٠,٩	٨,٢٣	٧,٠١	١٨
٣٦,٢	٣٢,٩	٣٠,١	٢٧,٢	٢٢,٧	١٨,٣	١٤,٦	١١,٧	٨,٩١	٧,٦٣	١٩
٣٧,٦	٣٤,٢	٣١,٤	٢٨,٤	٢٣,٨	١٩,٣	١٥,٥	١٢,٤	٩,٥٩	٨,٢٦	٢٠
٣٨,٩	٣٥,٥	٣٢,٧	٢٩,٦	٢٤,٩	٢٠,٣	١٦,٣	١٣,٢	١٠,٣	٨,٩٠	٢١
٤٠,٢	٣٦,٨	٣٣,٩	٣٠,٨	٢٦,٠	٢١,٣	١٧,٢	١٤,٥	١١,٠	٩,٥٤	٢٢
٤١,٦	٣٨,١	٣٥,٢	٣٢,٠	٢٧,١	٢٢,٣	١٨,١	١٤,٨	١١,٧	١٠,٢	٢٣
٤٣,٠	٣٩,٤	٣٦,٤	٣٣,٢	٢٨,٢	٢٣,٣	١٩,٠	١٥,٧	١٢,٤	١٠,٩	٢٤
٤٤,٢	٤٠,٦	٣٧,٧	٣٤,٤	٢٩,٣	٢٤,٣	١٩,٩	١٦,٥	١٣,١	١١,٥	٢٥
٤٥,٦	٤١,٩	٣٨,٩	٣٥,٦	٣٠,٤	٢٥,٣	٢٠,٨	١٧,٣	١٣,٨	١٢,٢	٢٦
٤٧,٠	٤٣,٢	٤٠,١	٣٦,٧	٣١,٥	٢٦,٣	٢١,٧	١٨,١	١٤,٦	١٢,٩	٢٧
٤٨,٢	٤٤,٥	٤١,٣	٣٧,٩	٣٢,٦	٢٧,٣	٢٢,٧	١٨,٩	١٥,٣	١٣,٦	٢٨
٤٩,٦	٤٥,٧	٤٢,٦	٣٩,١	٣٣,٧	٢٨,٣	٢٣,٦	١٩,٨	١٦,٠	١٤,٣	٢٩
٥٠,٩	٤٧,٠	٤٣,٨	٤٠,٢	٣٤,٨	٢٩,٣	٢٤,٥	٢٠,٦	١٦,٨	١٥,٠	٣٠
٥٢,٧	٤٩,٢	٤٥,٨	٤١,٨	٤٥,٦	٣٠,٣	٢٥,٧	٢١,٦	١٧,٤	١٥,٢	٤٠
٥٦,٢	٥١,٤	٤٧,٥	٤٣,٢	٤٩,٦	٣٤,٣	٢٧,٧	٢٣,٤	١٩,٧	١٧,٧	٥٠
٥٨,٤	٥٣,٣	٤٩,١	٤٥,٦	٥١,٣	٣٦,٣	٢٩,٥	٢٥,٥	٢١,٥	١٩,٥	٦٠

جدول رقم (٢ - ١١)
القيم الحرجة لاختبار كولموغوروف - سميرويف " د "
(٢٠٠٠ - ٢٠٠١)

درجات الحرجية	مستوى الدلالة				
	٠.١	٠.٠٥	٠.٠١	٠.٠٠٥	٠.٠٠١
١	٠.٩٩٥	٠.٩٧٥	٠.٩٥٠	٠.٩٢٥	٠.٩٠٠
٢	٠.٩٤٩	٠.٩٤٤	٠.٩٣٦	٠.٩٢٦	٠.٩١٤
٣	٠.٩٠٩	٠.٩٠٣	٠.٨٩٤	٠.٨٨٧	٠.٨٨٠
٤	٠.٨٧٤	٠.٨٦٤	٠.٨٥٦	٠.٨٤٦	٠.٨٣٩
٥	٠.٨٣٦	٠.٨٢٣	٠.٨١٠	٠.٨٠٤	٠.٧٩٦
٦	٠.٨١٨	٠.٨٠١	٠.٧٨٧	٠.٧٨٠	٠.٧٧١
٧	٠.٨٠٧	٠.٧٩٦	٠.٧٨٠	٠.٧٧٠	٠.٧٦١
٨	٠.٧٩٦	٠.٧٨٤	٠.٧٦٨	٠.٧٦٠	٠.٧٥١
٩	٠.٧٨٤	٠.٧٧٢	٠.٧٥٦	٠.٧٤٨	٠.٧٣٩
١٠	٠.٧٧٢	٠.٧٦٠	٠.٧٤٤	٠.٧٣٦	٠.٧٢٧
١١	٠.٧٦٠	٠.٧٤٨	٠.٧٣٢	٠.٧٢٤	٠.٧١٥
١٢	٠.٧٤٨	٠.٧٣٦	٠.٧٢٠	٠.٧١٢	٠.٧٠٣
١٣	٠.٧٣٦	٠.٧٢٤	٠.٧٠٨	٠.٧٠٠	٠.٦٩١
١٤	٠.٧٢٤	٠.٧١٢	٠.٦٩٦	٠.٦٨٨	٠.٦٧٩
١٥	٠.٧١٢	٠.٧٠٠	٠.٦٨٤	٠.٦٧٦	٠.٦٦٧
١٦	٠.٦٩٦	٠.٦٨٤	٠.٦٦٨	٠.٦٦٠	٠.٦٥١
١٧	٠.٦٨٤	٠.٦٧٢	٠.٦٥٦	٠.٦٤٨	٠.٦٣٩
١٨	٠.٦٧٢	٠.٦٦٠	٠.٦٤٤	٠.٦٣٦	٠.٦٢٧
١٩	٠.٦٦٠	٠.٦٤٨	٠.٦٣٢	٠.٦٢٤	٠.٦١٥
٢٠	٠.٦٤٨	٠.٦٣٦	٠.٦٢٠	٠.٦١٢	٠.٦٠٣
٢٥	٠.٦٠٣	٠.٥٩١	٠.٥٧٦	٠.٥٦٨	٠.٥٥٩
٣٠	٠.٥٥٩	٠.٥٤٧	٠.٥٣٢	٠.٥٢٤	٠.٥١٥
٤٠	٠.٥١٥	٠.٥٠٣	٠.٤٨٨	٠.٤٨٠	٠.٤٧١
٥٠	٠.٤٧١	٠.٤٥٩	٠.٤٤٤	٠.٤٣٦	٠.٤٢٧
٦٠	٠.٤٢٧	٠.٤١٥	٠.٣٩٦	٠.٣٨٨	٠.٣٧٩
٧٠	٠.٣٧٩	٠.٣٦٧	٠.٣٥٢	٠.٣٤٤	٠.٣٣٥
٨٠	٠.٣٣٥	٠.٣٢٣	٠.٣٠٨	٠.٣٠٠	٠.٢٩١
٩٠	٠.٢٩١	٠.٢٧٩	٠.٢٦٤	٠.٢٥٦	٠.٢٤٧
١٠٠	٠.٢٤٧	٠.٢٣٥	٠.٢٢٠	٠.٢١٢	٠.٢٠٣
المتوسط	٠.٢٠٣	٠.١٩١	٠.١٧٦	٠.١٦٨	٠.١٥٩

(يرفض فرض العدم عندما تكون قيمة " د " أكبر من القيمة الحرجية عند مستوى الدلالة المطلوبه) .

جدول رقم (م-١٥)
البيانات المرجحة لاختبار هامان - هوبنثي (اختبار هـ) عند مستوى دلالة (٠.٠٥) اختبار تكي الطريق

١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	١١	١٢	١٣	١٤	١٥	١٦	١٧	١٨	١٩	٢٠
١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢
٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣
٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤
٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥
٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦
٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧
٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨
٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩
١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠
١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١
١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢
١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣
١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤
١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥
١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦
١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧
١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨
١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩
٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠

يوضح فرض الدم عند مستوى دلالة ٠.٠٥ في السوية أفقصة أولساوي القيمة المرجحة عند مستوى دلالة ٠.٠٥

جدول رسم (١٢-٢)
(تابع) حيدو رسم (اختيارية) عند مستوى دلالة ٠.٠٥ [مبتدئ في المرحلة]

ن.م	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	١١	١٢	١٣	١٤	١٥	١٦	١٧	١٨	١٩	٢٠
١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٢	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٣	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٤	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٥	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٦	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٧	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٨	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٩	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
١٠	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
١١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
١٢	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
١٣	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
١٤	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
١٥	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
١٦	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
١٧	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
١٨	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
١٩	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٢٠	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١

الفر المرحبة لاختبار مان - هويقي (اختبار - ي) عند مستوى دلالة ٥.٠ م [اختبار شافلي الطرف]
 (نتاج) جدول رتبه (٣ - ١٩)

رتبه	١	٢	٣	٤	٥	٦	٧	٨	٩	١٠	١١	١٢	١٣	١٤	١٥	١٦	١٧	١٨	١٩	٢٠
١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١	١
٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢	٢
٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣	٣
٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤	٤
٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥	٥
٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦	٦
٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧	٧
٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨	٨
٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩	٩
١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠	١٠
١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١	١١
١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢	١٢
١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣	١٣
١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤	١٤
١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥	١٥
١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦	١٦
١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧	١٧
١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨	١٨
١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩	١٩
٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠	٢٠

جدول رقم (٢-١٢)
القيم المرجحة لاختبار ويلكوسون (اختبار -٥)

عدد أزواج القيم = ن	١ - ٥٠		١ - ٥٠ ٥ - ٢٥
	١ - ٥٠	١ - ٥٠	
٦	١	١	١
٧	٢	٢	٢
٨	٣	٣	٣
٩	٤	٤	٤
١٠	٥	٥	٥
١١	٦	٦	٦
١٢	٧	٧	٧
١٣	٨	٨	٨
١٤	٩	٩	٩
١٥	١٠	١٠	١٠
١٦	١١	١١	١١
١٧	١٢	١٢	١٢
١٨	١٣	١٣	١٣
١٩	١٤	١٤	١٤
٢٠	١٥	١٥	١٥
٢١	١٦	١٦	١٦
٢٢	١٧	١٧	١٧
٢٣	١٨	١٨	١٨
٢٤	١٩	١٩	١٩
٢٥	٢٠	٢٠	٢٠
٢٦	٢١	٢١	٢١
٢٧	٢٢	٢٢	٢٢
٢٨	٢٣	٢٣	٢٣
٢٩	٢٤	٢٤	٢٤
٣٠	٢٥	٢٥	٢٥
٣١	٢٦	٢٦	٢٦
٣٢	٢٧	٢٧	٢٧
٣٣	٢٨	٢٨	٢٨
٣٤	٢٩	٢٩	٢٩
٣٥	٣٠	٣٠	٣٠
٣٦	٣١	٣١	٣١
٣٧	٣٢	٣٢	٣٢
٣٨	٣٣	٣٣	٣٣
٣٩	٣٤	٣٤	٣٤
٤٠	٣٥	٣٥	٣٥
٤١	٣٦	٣٦	٣٦
٤٢	٣٧	٣٧	٣٧
٤٣	٣٨	٣٨	٣٨
٤٤	٣٩	٣٩	٣٩
٤٥	٤٠	٤٠	٤٠
٤٦	٤١	٤١	٤١
٤٧	٤٢	٤٢	٤٢
٤٨	٤٣	٤٣	٤٣
٤٩	٤٤	٤٤	٤٤
٥٠	٤٥	٤٥	٤٥

١ - اختبار ثنائي الطرفين
٢ - اختبار أحادي الطرفين

(قيمة ٥ - المسيرة تكون ذات دلالة إحصائية إذا كانت أقل من القيمة
النظرية لـ ٥ في الجدول عند مستوى الدلالة المطلوب).

جدول رقم (٢ - ١٤)
القيم المرجحة لاختبار كروسكال - ونيس
اختبار هـ.

مستوى الدلالة				نم نم نم		
٥٠٠	١٠٠	٥٠	١٠	١	٢	٣
			٤,٥٧١	١	٢	٣
			٤,٥٨٦	١	٢	٣
			٤,٥٠٠	١	٢	٣
	٥,٢٥٧	٤,٧١٤	٤,٥٧١	١	٢	٣
		٥,١٤٣	٤,٥٧١	١	٢	٣
		٥,٣٦١	٤,٥٥٦	١	٢	٣
٤,٤٠٠	٦,٤٠٠	٥,٦٠٠	٤,٦٤٤	٢	٣	٤
			٤,٥٠٠	١	٢	٤
			٤,٥٠٦	١	٢	٤
		٥,٢٠٨	٤,٥١١	١	٢	٤
	٥,٤٤٤	٥,٤٤٤	٤,٧٠٩	٢	٣	٤
	٥,٤٦٦	٥,٧٤٧	٤,١٦٧	١	٤	٤
	٥,٦٦٧	٤,٨٦٧	٤,٥٥٥	١	٤	٤
	٥,٦٦٧	٥,٤٥٥	٤,٥٤٦	٢	٤	٤
	٥,٦٦٧	٥,٥٩٩	٤,٦٥٤	٢	٤	٤
	٥,٦٦٧	٥,٦٩٤	٤,٦٥٤	٢	٤	٤
		٥,٠٠٠	٤,٦٠٠	١	٤	٥
	٥,٢٢٢	٤,٦١٠	٤,٦٧٢	١	٤	٥
		٤,٩٦٠	٤,٦١٨	١	٤	٥
	٥,٢٨٤	٥,٢٥١	٤,٦٥١	١	٤	٥
	٥,٢٨٤	٤,٦٤٩	٤,٥٢٢	٢	٤	٥
	٥,٢٨٤	٤,٦٨٦	٤,٦٨٧	٢	٤	٥
	٥,٢٨٤	٥,٢٦٨	٤,٥٢١	٢	٤	٥
	٥,٢٨٤	٥,٢٣١	٤,٥٤٩	٢	٤	٥
	٥,٢٨٤	٥,٢١٨	٤,٦١٩	٢	٤	٥
	٥,٢٨٤	٥,١٤٧	٤,٦٠٩	٢	٤	٥
	٥,٢٨٤	٥,٢٢٩	٤,٥٠٨	٢	٤	٥
	٥,٢٨٤	٥,٢٠٦	٤,٥٤٥	٢	٤	٥
	٥,٢٨٤	٥,٢٤٣	٤,٥٢٢	٢	٤	٥
	٥,٢٨٤	٥,٢٨٠	٤,٥٦٠	٢	٤	٥

(يرفض فرض العدم عندما تكون قيمة هـ أكبر من دالة الاختبار)
القيمة الحرجة عند مستوى الدلالة المطلوب).

جدول رقم (١٥٣) : قيمة أقل مماثل ارتباط معنوي عند مستوى دلالة ٥.٠ و ٥.٠٥

درجات الحرية	٥	١	درجات الحرية	٥	١
١	٠,٩٩٧	١,٥٠٠	٢٤	٠,٢٨٨	٠,٤٩٦
٢	٠,٩٥٠	٠,٩٩٠	٢٥	٠,٢٨١	٠,٤٨٧
٣	٠,٨٧٨	٠,٩٥٩	٢٦	٠,٢٧٤	٠,٤٧٨
٤	٠,٨١١	٠,٩١٧	٢٧	٠,٢٦٧	٠,٤٧٠
٥	٠,٧٥٤	٠,٨٧٤	٢٨	٠,٢٦١	٠,٤٦٣
٦	٠,٧٠٧	٠,٨٣٤	٢٩	٠,٢٥٥	٠,٤٥٦
٧	٠,٦٦٦	٠,٧٩٨	٣٠	٠,٢٤٩	٠,٤٤٩
٨	٠,٦٢٢	٠,٧٦٥	٣٥	٠,٢٣٥	٠,٤١٨
٩	٠,٦٠٢	٠,٧٣٥	٤٠	٠,٢٢٤	٠,٣٩٣
١٠	٠,٥٨٦	٠,٧٠٨	٤٥	٠,٢١٨	٠,٣٧٢
١١	٠,٥٧٣	٠,٦٨٤	٥٠	٠,٢١٢	٠,٣٦٤
١٢	٠,٥٦٣	٠,٦٦١	٦٠	٠,٢٠٥	٠,٣٣٥
١٣	٠,٥٥٤	٠,٦٤١	٧٠	٠,٢٠٢	٠,٣٢٢
١٤	٠,٥٤٧	٠,٦٢٣	٨٠	٠,١٩٧	٠,٢٨٣
١٥	٠,٥٤٢	٠,٦٠٦	٩٠	٠,١٩٠	٠,٢٦٧
١٦	٠,٥٣٨	٠,٥٩٠	١٠٠	٠,١٨٥	٠,٢٥٤
١٧	٠,٥٣٦	٠,٥٧٥	١٢٥	٠,١٧٤	٠,٢٢٨
١٨	٠,٥٣٤	٠,٥٦١	١٥٠	٠,١٥٩	٠,٢٠٨
١٩	٠,٥٣٣	٠,٥٤٩	٢٠٠	٠,١٣٨	٠,١٨١
٢٠	٠,٥٣٢	٠,٥٣٧	٢٥٠	٠,١١٣	٠,١٤٨
٢١	٠,٥٣١	٠,٥٢٦	٣٠٠	٠,٠٩٨	٠,١٢٨
٢٢	٠,٥٣٠	٠,٥١٥	٤٠٠	٠,٠٨٨	٠,١١٥
٢٣	٠,٥٢٩	٠,٥٠٥	١٠٠٠	٠,٠٦٢	٠,٠٨١

جدول ٢١-١٦)
قيمة أقل معامل ارتباط معنوي لمعامل ارتباط الرتب (سبيرمان)

الدرجة المرتبة	١-٥	٥-١٠	١٠-٢٠	٢٠-٥٠
١	١٠٠٠	١٠٠٠	١٠٠٠	١٠٠٠
٢	٨٩٩	٨٩٩	٨٩٩	٨٩٩
٣	٨٨٨	٨٨٨	٨٨٨	٨٨٨
٤	٨٧٧	٨٧٧	٨٧٧	٨٧٧
٥	٨٦٦	٨٦٦	٨٦٦	٨٦٦
٦	٨٥٥	٨٥٥	٨٥٥	٨٥٥
٧	٨٤٤	٨٤٤	٨٤٤	٨٤٤
٨	٨٣٣	٨٣٣	٨٣٣	٨٣٣
٩	٨٢٢	٨٢٢	٨٢٢	٨٢٢
١٠	٨١١	٨١١	٨١١	٨١١
١١	٨٠٠	٨٠٠	٨٠٠	٨٠٠
١٢	٧٩٩	٧٩٩	٧٩٩	٧٩٩
١٣	٧٨٨	٧٨٨	٧٨٨	٧٨٨
١٤	٧٧٧	٧٧٧	٧٧٧	٧٧٧
١٥	٧٦٦	٧٦٦	٧٦٦	٧٦٦
١٦	٧٥٥	٧٥٥	٧٥٥	٧٥٥
١٧	٧٤٤	٧٤٤	٧٤٤	٧٤٤
١٨	٧٣٣	٧٣٣	٧٣٣	٧٣٣
١٩	٧٢٢	٧٢٢	٧٢٢	٧٢٢
٢٠	٧١١	٧١١	٧١١	٧١١
٢١	٧٠٠	٧٠٠	٧٠٠	٧٠٠
٢٢	٦٩٩	٦٩٩	٦٩٩	٦٩٩
٢٣	٦٨٨	٦٨٨	٦٨٨	٦٨٨
٢٤	٦٧٧	٦٧٧	٦٧٧	٦٧٧
٢٥	٦٦٦	٦٦٦	٦٦٦	٦٦٦
٢٦	٦٥٥	٦٥٥	٦٥٥	٦٥٥
٢٧	٦٤٤	٦٤٤	٦٤٤	٦٤٤
٢٨	٦٣٣	٦٣٣	٦٣٣	٦٣٣
٢٩	٦٢٢	٦٢٢	٦٢٢	٦٢٢
٣٠	٦١١	٦١١	٦١١	٦١١
٣١	٦٠٠	٦٠٠	٦٠٠	٦٠٠
٣٢	٥٩٩	٥٩٩	٥٩٩	٥٩٩
٣٣	٥٨٨	٥٨٨	٥٨٨	٥٨٨
٣٤	٥٧٧	٥٧٧	٥٧٧	٥٧٧
٣٥	٥٦٦	٥٦٦	٥٦٦	٥٦٦
٣٦	٥٥٥	٥٥٥	٥٥٥	٥٥٥
٣٧	٥٤٤	٥٤٤	٥٤٤	٥٤٤
٣٨	٥٣٣	٥٣٣	٥٣٣	٥٣٣
٣٩	٥٢٢	٥٢٢	٥٢٢	٥٢٢
٤٠	٥١١	٥١١	٥١١	٥١١
٤١	٥٠٠	٥٠٠	٥٠٠	٥٠٠
٤٢	٤٩٩	٤٩٩	٤٩٩	٤٩٩
٤٣	٤٨٨	٤٨٨	٤٨٨	٤٨٨
٤٤	٤٧٧	٤٧٧	٤٧٧	٤٧٧
٤٥	٤٦٦	٤٦٦	٤٦٦	٤٦٦
٤٦	٤٥٥	٤٥٥	٤٥٥	٤٥٥
٤٧	٤٤٤	٤٤٤	٤٤٤	٤٤٤
٤٨	٤٣٣	٤٣٣	٤٣٣	٤٣٣
٤٩	٤٢٢	٤٢٢	٤٢٢	٤٢٢
٥٠	٤١١	٤١١	٤١١	٤١١
٥١	٤٠٠	٤٠٠	٤٠٠	٤٠٠
٥٢	٣٩٩	٣٩٩	٣٩٩	٣٩٩
٥٣	٣٨٨	٣٨٨	٣٨٨	٣٨٨
٥٤	٣٧٧	٣٧٧	٣٧٧	٣٧٧
٥٥	٣٦٦	٣٦٦	٣٦٦	٣٦٦
٥٦	٣٥٥	٣٥٥	٣٥٥	٣٥٥
٥٧	٣٤٤	٣٤٤	٣٤٤	٣٤٤
٥٨	٣٣٣	٣٣٣	٣٣٣	٣٣٣
٥٩	٣٢٢	٣٢٢	٣٢٢	٣٢٢
٦٠	٣١١	٣١١	٣١١	٣١١
٦١	٣٠٠	٣٠٠	٣٠٠	٣٠٠
٦٢	٢٩٩	٢٩٩	٢٩٩	٢٩٩
٦٣	٢٨٨	٢٨٨	٢٨٨	٢٨٨
٦٤	٢٧٧	٢٧٧	٢٧٧	٢٧٧
٦٥	٢٦٦	٢٦٦	٢٦٦	٢٦٦
٦٦	٢٥٥	٢٥٥	٢٥٥	٢٥٥
٦٧	٢٤٤	٢٤٤	٢٤٤	٢٤٤
٦٨	٢٣٣	٢٣٣	٢٣٣	٢٣٣
٦٩	٢٢٢	٢٢٢	٢٢٢	٢٢٢
٧٠	٢١١	٢١١	٢١١	٢١١
٧١	٢٠٠	٢٠٠	٢٠٠	٢٠٠
٧٢	١٩٩	١٩٩	١٩٩	١٩٩
٧٣	١٨٨	١٨٨	١٨٨	١٨٨
٧٤	١٧٧	١٧٧	١٧٧	١٧٧
٧٥	١٦٦	١٦٦	١٦٦	١٦٦
٧٦	١٥٥	١٥٥	١٥٥	١٥٥
٧٧	١٤٤	١٤٤	١٤٤	١٤٤
٧٨	١٣٣	١٣٣	١٣٣	١٣٣
٧٩	١٢٢	١٢٢	١٢٢	١٢٢
٨٠	١١١	١١١	١١١	١١١
٨١	١٠٠	١٠٠	١٠٠	١٠٠

١- اختبار تناك المربعات
٢- اختبار أحادي الطرفين

